Structural and functional characterization of adhesins and the virulence regulator RovC of *Yersinia pseudotuberculosis*

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades einer
Doktorin der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Pooja Sadana

aus Saharanpur, India

1. Referent:	Prof. Dr. Wulf Blankenfeldt
2. Referent:	Prof. Dr. Stefan Dübel
3. Referent:	Prof. Dr. Hartmut Niemann
eingereicht am:	25.01.2017
mündliche Prüfung (Disputation) am:	01.06.2017

Druckjahr 2017

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Sadana, P., Geyer, R., Dersch, P. & Scrima, A. Structural and functional studies of Invasin-like proteins. (Poster). 5th Annual Retreat- Helmholtz Centre for Infection Research Graduate School, Goslar Hahnenklee, Germany (May 2014).

Sadana, P., Geyer, R., Dersch, P. & Scrima, A. Structural and functional studies of Invasin-like proteins. (Vortrag). 17th Heart-of-Europe Bio-Crystallography meeting, Berlin, Germany (September 2014).

Sadana, P., Geyer, R., Dersch, P. & Scrima, A. Structural and functional studies of Invasin-like proteins. (Poster). North Regio Day on Infection (NoRDI) V, Helmholtz Centre for Infection Research, Braunschweig, Germany (October 2014).

Sadana, P., Geyer, R., Dersch, P. & Scrima, A. Structural and functional studies of Invasin-like protein. (Poster). 7th International PhD symposium of the Helmholtz International Graduate School for Infection Research, Braunschweig, Germany (December 2014).

Sadana, P., Geyer, R., Dersch, P. & Scrima, A. Structural and functional studies of Invasin-like proteins. (Poster). 23rd Annual Conference of the German Crystallographic society, Göttingen, Germany (March 2015).

Sadana, P., Dersch, P. & Scrima, A. Structural and functional studies of Invasin-like protein. (Poster). 8th International PhD symposium of the Helmholtz International Graduate School for Infection Research, Braunschweig, Germany (December 2015).

Sadana, P., Geyer, R., Pezoldt, J., Helmsing, S., Huehn, J., Hust, M., Dersch, P., & Scrima, A. Structural and functional characterization of *Yersinia pesuotuberculosis* adhesin InvD.

(Vortrag). 19th Heart-of-Europe Bio-Crystallography meeting, Burg Warberg, Germany (Sepetmber 2016).

Sadana, P., Geyer, R., Pezoldt, J., Helmsing, S., Huehn, J., Hust, M., Dersch, P. & Scrima, A. Structural and functional characterization of *Yersinia pesuotuberculosis* adhesin InvD. (Vortrag). 9th International PhD symposium of the Helmholtz International Graduate School, Helmholtz Centre for Infection Research, Braunschweig, Germany (December 2016).

Contents

Summary

Y. pseudotuberculosis is a human pathogenic bacterium that is responsible for several gastrointestinal diseases commonly called yersiniosis. The pathogen is generally ingested as contaminated food or water. Once in the intestine, *Yersinia* enters the lymphatic system through the M-cells of the small intestine to colonize and invade their hosts.

Yersinia has several multifunctional adhesins on its surface that contribute to its pathogenicity. These virulence factors are in turn controlled by complex regulatory networks that control the expression of these genes, which help *Yersinia* to survive and adapt to their rapidly changing environment. The projects presented here involve the structural and functional characterization of adhesins and the virulence regulator RovC of *Y. pseudotuberculosis.*

Enteropathogenic *Yersiniae* express three important adhesins, belonging to different protein classes: Invasin, YadA and Ail. Invasin-like proteins represent a subgroup of the adhesin family that specifically interacts with receptors on the host cell plasma membrane. The C-terminal domain of Invasins provides specificity for interaction with host-derived factors. Invasin (InvA) is the major adhesion factor of *Y. pseudotuberculosis* and it promotes binding of *Yersinia* to epithelial cells. In addition to InvA, three additional Invasins InvB, InvC and InvD have been identified in *Y. pseudotuberculosis* YPIII. Unlike InvB and InvC whose C-terminal domain share high sequence similarity with InvA and thus are likely to share a common fold, InvD C-terminal domain shares no sequence homology with any of the other three Invasins. The aim of this work was to get insight into the structure and function of the C-terminal domain of InvD.

To achieve this, InvD construct with two Big-like domains and the adhesion domain was cloned, expressed, purified and crystallized. The structure was solved to 2.6 Å resolution by combining molecular replacement with single wavelength anomalous dispersion (MR-SAD). The overall shape seen in the crystal structure was confirmed in solution by Small Angle X-ray Scattering (SAXS). The N-terminal domains of crystallized InvD reveal similarities to the Immunoglobulin Superfamily (IgSF) fold. In contrast, the C-terminal domain adopts a novel fold. The surface of the adhesion domain clearly shows two highly conserved patches that might be involved in targeting the host molecule. Fab fragment of immunoglobulins was identified as the host interaction partner of InvD by pull downs and microscale thermophoresis (MST) experiments. Phage display panning, pull downs and MST experiments confirmed the specificity of InvD towards VH3/VK1 fragment of the antibody. Furthermore, fluorescent activated cell sorting experiments revealed that InvD binds specifically to B-cell population while no binding was observed with the other immune cells.

In addition to InvD, a novel Invasin, InvE has been identified in the IP31758 strain of *Y. pseudotuberculosis*. First 1730 aa of InvE are highly homologous to that of InvD, but the C-terminal domain is significantly different. Hence, the aim of this project was to gain insights into the structure of InvE. To achieve this, InvE construct with two BIg-like domains and the adhesion domain was cloned, expressed, purified and crystallized. The crystals obtained diffracted to approx. 8 Å resolution. Any attempts to improve the crystal quality were not successful. Therefore, surface entropy reduction (SER) approach was used to alter crystallization behavior. The crystals obtained with SER mutant diffracted to 1.7 Å and the structure was solved by using a Seleno-Methionine (Se-Met) variant of InvE SER mutant. The rod shaped structure of InvE was confirmed in solution by SAXS. The N-terminal domains of crystallized InvE reveal similarities to the IgSF fold. In contrast, the C-terminal domain adopts a C-type lectin like fold and share high structural similarity with the C-type lectin like domain of InvA and Intimin, even though there exists no sequence homology among these proteins.

YadA of *Y. enterocolitica* (ent) and *Y. pseudotuberculosis* (pstb) exhibit differences in their specificity of extracellular matrix (ECM) substrate binding. This difference is due to a unique N-terminal 31 aa sequence which is present in $YadA_{pstb}$ and absent in $YadA_{ent}$. The goal of this work was to provide the structural basis for the functional transition by solving the structure of $YadA_{pstb}$ and to compare it with the known $YadA_{ent}$ structure. For that purpose, the head region of $YadA_{pstb}$ was cloned, expressed, purified and crystallized. The structure was solved to 2.3 Å resolution by molecular replacement. However, due to the high flexibility, this insertion was only partially visible in the crystal structure. Despite of the partial flexibility, the results show that the insertion region is located at the tip, which is likely to change the tip surface and influence ECM specificity.

A novel virulence regulator, RovC has been recently identified in *Y. pseudotuberculosis* that activates Type6 secretion system-4 transcription. RovC is highly conserved among *Y. pestis* and *Y. pseudotuberculosis*, but does not exhibit homology to any other known protein. The aim of this work was to get insights into the structure and DNA binding activity of this novel regulator. For that purpose, full length RovC was cloned, expressed, purified and crystallized. The structure was solved to 2.3 Å resolution by using a Se-Met variant of RovC. Crystallized RovC consists of two domains- the N-terminal domain comprised of α-helices and β-strands that adopts a distinct fold, and the C-terminal domain containing a helix-turn-helix motif. SAXS analysis of RovC revealed the hexameric ring-shaped structure which was validated by multi angle light scattering and site-directed mutagenesis experiments. Furthermore, site-directed mutagenesis approach and MST were used to characterize the DNA binding activity and the model has been proposed in which DNA wraps around the surface of hexameric RovC.

1 Introduction

In order to survive in their niche, bacterial pathogens armor themselves with a variety of virulence factors that are needed by them to adhere and invade the host tissues, escape and modulate the host immune system. The expression of these virulence factors is in turn controlled by complex regulatory networks. This helps bacteria to survive and adapt to their rapidly changing environment (Erhardt and Dersch, 2015).

1.1 The genus *Yersinia*

The genus *Yersinia* was named after a swiss-born French physician Alexandre Jean Emile Yersin (1863-1943) who discovered the plague bacillus in 1894 (Hawgood, 2008).

The genus *Yersinia* comprises seventeen different species of which *Yersinia pestis, Yersinia pseudotuberculosis* and *Yersinia enterocolitica* are pathogenic to humans (Savin et al., 2014). All these species show high tropism for lymphatic tissues and the ability to protect themselves against the host immune system (Balada-Llasat and Mecsas, 2006).

Y. pestis is commonly transmitted to humans by the bites of infected fleas and causes plague, while *Y. pseudotuberculosis* and *Y. enterocolitica* are enteropathogenic to humans. *Y. pseudotuberculosis* and *Y. pestis* have a nearly identical chromosomal DNA sequence. *Y. pestis* genome shares 83% homology to *Y. pseudotuberculosis* (Moore and Brubaker, 1975) and it is now evident that *Y. pestis* has evolved from *Y. pseudotuberculosis* (Achtman et al., 1999). Although the life cycle and the character of disease caused by *Y. pestis* is very different from that of *Y. pseudotuberculosis*, these two species are so closely related that *Y. pestis* is classified as a *pseudotuberculosis* subspecies.

1.2 Mode of infection of *Y. pestis*

Y. pestis is a causative agent for a systemic infectious disease known as plague. The major carriers of *Y. pestis* are ground dwelling rodents which transmit disease to humans through infected flea bites (Figure 1). *Y. pestis* then resides in the lymphatic tissues, which leads to bubonic plague. This is the most commonly occurring form of plague. In some cases, inhaling *Y. pestis* particles present in the air can lead to pneumonic plague which affects the lungs. Pneumonic plague is the most uncommon, but the most deadly form of plague. In rare cases, when *Y. pestis* multiplies in the blood stream, it can also lead to septicemic plague (Heroven and Dersch, 2014).

When *Y. pestis* invades the host skin, it encounters macrophages and other immune cells at the site of invasion. To escape the host immune system, it survives and replicates within the host macrophages during the early stages of infection. This intracellular growth is necessary for the pathogenesis of bacteria. When the bacterium reaches the lymph nodes, it escapes

from the macrophages, develops resistance to phagocytosis and survives extracellularly. At this point, the bacteria use several Type 3 Secretion System (T3SS) effector proteins to inject into the target cell, to further modulate the host immune system (Li and Yang, 2008; Mikula et al., 2012).

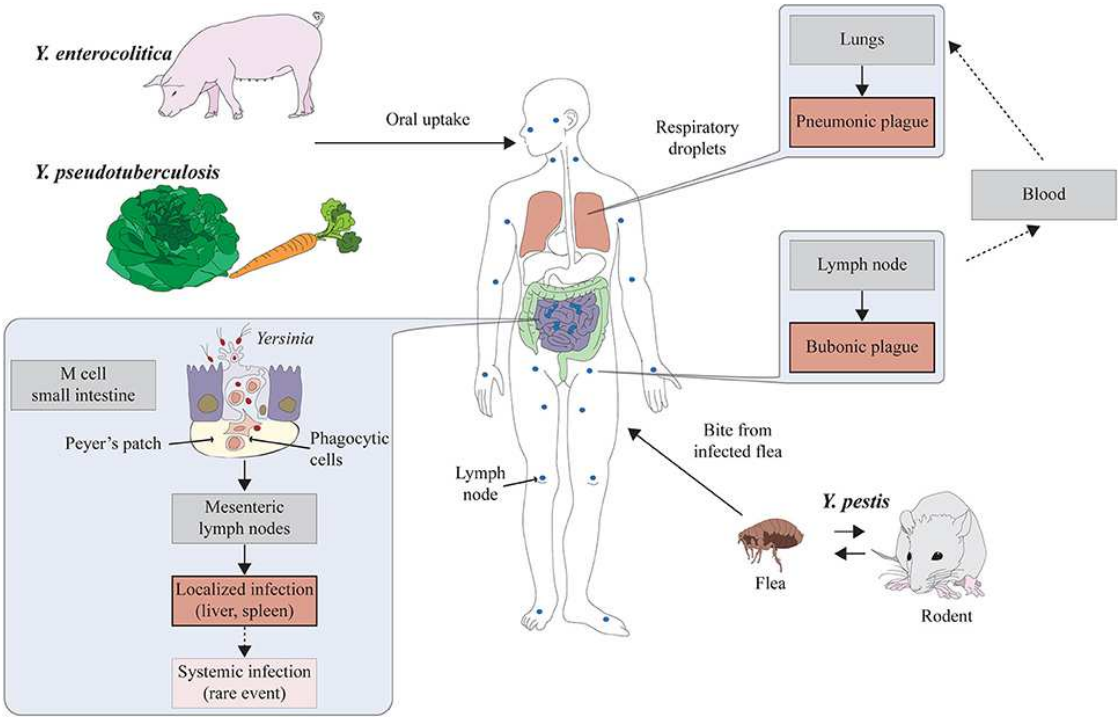

Figure 1: Infection routes of pathogenic *Yersinia*. The main reservoirs for *Y. pseudotuberculosis* and *Y. enterocolitica* are pork and fresh vegetables. Once ingested, the bacteria cross the epithelial barrier and enter the lymphatic system. In contrast to the two enteropathogenic *Yersinia* species, *Y. pestis* enters the human lymphatic system by the bite of an infected flea and causes bubonic plague (Adapted from (Heroven and Dersch, 2014)).

1.3 Mode of infection of *Y. enterocolitica* and *Y. pseudotuberculosis*

Y. enterocolitica and *Y. pseudotuberculosis* are food-borne pathogens that cause several gastrointestinal diseases (yersinioisis) which further lead to enteritis, acute mesenteric lymphadenitis and terminal ileitis. Both of these species exhibit 55% sequence identity at the nucleotide level (Snellings et al., 2001). For *Y. enterocolitica*, swine is the most common reservoir from which pathogenic strains have regularly been isolated (Bottone, 1999). *Y. pseudotuberculosis* pathogenic strains are on the other hand isolated from different sources such as water, soil, fresh vegetables and wild animals (Heroven and Dersch, 2014). These are generally ingested as contaminated food or water and traverse the gastrointestinal tract (Figure 1).

Once *Yersinia* reaches the small intestine, it encounters microfold cells (M cells). M cells are specialized epithelial cells of the follicle-associated epithelium (FAE) of the gastrointestinal tract and are responsible for the uptake of antigens and microorganisms and delivery to the

immune cells of the lymphoid tissues (Kucharzik et al., 2000). To invade epithelial cells, *Yersinia* binds to at least five different β_1- integrin receptors ($\alpha_3\beta_1$, $\alpha_4\beta_1$, $\alpha_5\beta_1$, $\alpha_6\beta_1$ and $\alpha_v\beta_1$) expressed on the surface of M cells (Isberg and Leong, 1990). The adhesion is mediated by the outer membrane protein Invasin (InvA), which is the very first virulence factor required by *Yersinia* during the early stages of infection. This interaction further induces the interaction of cytoplasmic β_1 chain with focal adhesion kinases (FAKs), Src and Rac-1-Arp2/3 complex which leads to cytoskeletal rearrangement, a prerequisite for bacterial uptake. This process is called zipper mechanism. *Yersinia* internalized by zipper mechanism, remains intracellular surrounded by the vacuole. The bacterium is released from the vacuole into the dome region of FAE. Once *Yersinia* has translocated across the epithelial barrier, it is exposed to a variety of immune cells such as dendritic cells, macrophages and lymphocytes. At this stage, *Yersinia* injects several Yop (*Yersinia* outer proteins) into the macrophages which provides resistance to bacteria against phagocytosis and help it to survive extracellularly. Furthermore, virulence factors of *Yersinia* like YadA (Yersinia adhesion A) and Ail (Adherence and Invasion Locus) provide resistance against complement mediated lysis (Barnes et al., 2006; Reis and Horn, 2010) (Figure 2).

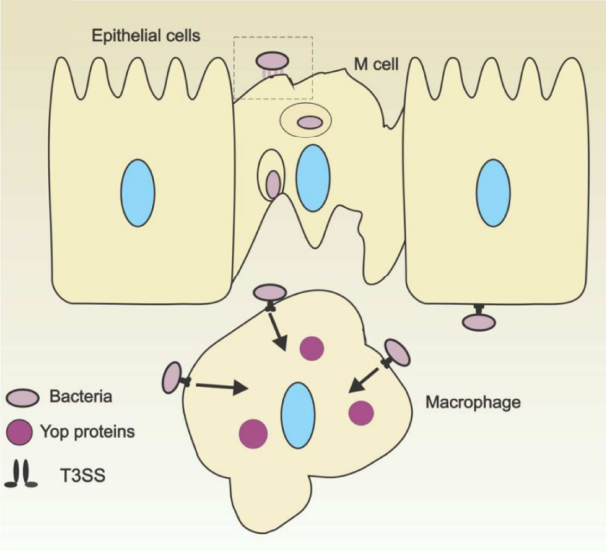

Figure 2: Route of infection of enteropathogenic *Yersinia*. Via consumption of contaminated food or water, the pathogen is ingested and reaches the small intestine. Once in the intestine, it binds to β_1- integrins present on the M cells, which resulted in the uptake of bacteria by zipper mechanism. Once the bacteria reach the dome region of FAE, it is exposed to a variety of immune cells. Thereby *Yersinia* uses several of its virulence factors such as (Yop and T3SS) to escape the host system (Adapted from (Reis and Horn, 2010)).

1.4 Virulence factors of *Yersinia*

To survive and replicate inside the host, *Yersinia* carries several essential plasmids and chromosome encoded virulence factors.

All the pathogenic strains of *Yersinia* carry a 70 kb virulence plasmid termed pYV (plasmid for *Yersinia* virulence) which differentiates pathogenic from non-pathogenic strains. The pYV plasmid is also called low calcium response plasmid (pLCR) as it restricts the growth of

Yersinia in calcium deficient media. This plasmid is essential for virulence in all the three species of *Yersinia* (Hu et al., 1998; Huang et al., 2006).

The pYV plasmid codes for YadA, T3SS and several Yops which are essential for bacterial replication in the host. YadA is a multifunctional surface exposed protein that binds to extracellular matrix proteins and also provides resistance to serum complement lysis. The pYV encoded T3SS injects different Yops into the host cell which interferes with the host signal transduction pathway, provides resistance to *Yersinia* against phagocytosis by macrophages and polymorphonuclear leukocytes and also promotes the survival of *Yersinia* inside the cell by blocking the inflammatory response of cells to infection (Cornelis, 2002).

Besides the plasmid associated pathogenicity factors, *Yersinia* also harbors *inv* and *ail* which are chromosomally encoded virulence genes. *inv* codes for Invasin, which is an essential protein required by *Yersinia* to get entry inside the cell during the initial stages of *Y. pseudotuberculosis* and *Y. enterocolitica* infection. *ail* codes for Ail protein, which plays a significant role in immune evasion (Biedzka-Sarek et al., 2008). In contrast to *inv*, *ail* is only found in *Y. enterocolitica*. Strains that lack the virulence plasmid remain invasive indicating that the factors responsible for invasion are located on the chromosome of *Yersinia* (Portnoy et al., 1981).

Pathogenic species of *Yersinia* also possess a large chromosomal fragment known as the High Pathogenicity Island (HPI) which encodes for yersiniabactin system and is involved in siderophore mediated iron-uptake (Perry and Fetherston, 2011).

Two specific plasmids that are unique to *Y. pestis*, and are not shared by the enteropathogenic species of *Yersinia* are pMT1/pFra and pPCP1/pPst. The 100 kb pMT1/pFra encodes for murine toxin and F1 capsular antigen, former is required for survival of *Y. pestis* in the flea gut, while the latter provides resistance against phagocytosis. The 9.6 kb plasmid pPCP1/pPst encodes for the Plasminogen activator (Pla) which is required by *Y. pestis* for efficient dissemination in host tissues (Du et al., 2002; Huang et al., 2006; Rajanna et al., 2010).

The uncharacterized plasmid pVM82 has been reported in the strain of *Y. pseudotuberculosis* IP31758. It could be shown that pVM82 provides antiphagocytic and immunosuppressive capabilities to *Y. pseudotuberculosis*. It also codes for type IVB icm/dot secretion system. This system is only present in IP31758 strain of *Y. pseudotuberculosis* and might mediate its intracellular survival in epithelial cells (Eppinger et al., 2007).

1.5 *Yersinia* adhesins

In order to infect a cell, *Yersinia* possesses multifunctional adhesins (Figure 3) on its surface. These are not only required by bacteria to get entry inside the cell but are also required for efficient colonization and infection.

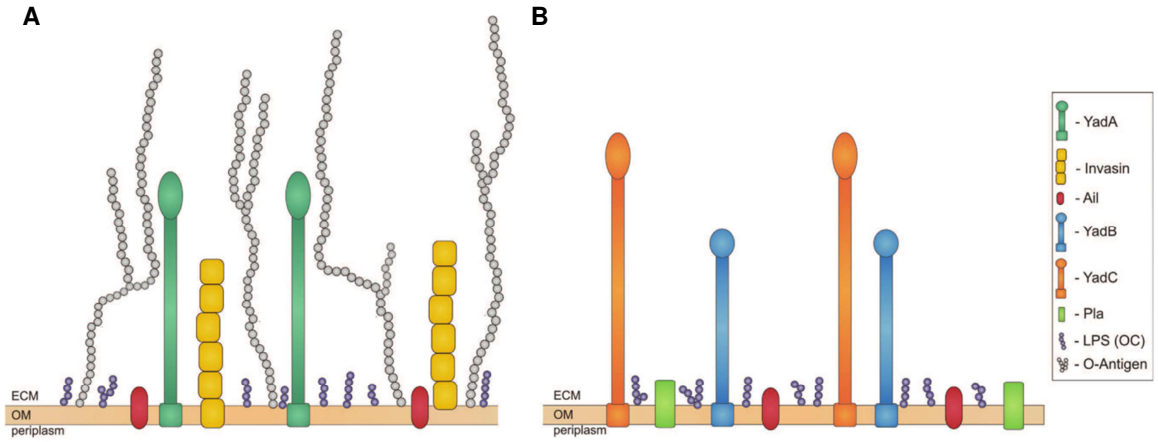

Figure 3: Schematic overview of different outer membrane proteins expressed by pathogenic *Yersinia.* A) Adhesins of *Y. pseudotuberculosis* and *Y. enterocolitica*. B) Adhesins of *Y. pestis*. ECM stands for extracellular matrix, OM stands for Outer membrane. (Adapted from (Mikula et al., 2012)).

1.5.1 Invasin-like proteins

Invasin or Invasin-like proteins of *Y. pseudotuberculosis* belong to the family of adhesins that share a common architecture (Figure 4A). They consist of i) an N-terminal highly homologous β-barrel-like domain, which is responsible for the attachment of Invasins to the outer membrane region of bacteria, ii) repetitive Ig-like domains, which vary significantly in number among Invasins and iii) typically a C-terminal C-type lectin domain, often described as the capping or adhesion domain, which provides specificity for interaction with host-derived factors (Pisano et al., 2012).

The arrangement of domains present in Invasins is similar to that found in classical autotransporters (Va). However, the order of domains is reversed in Invasins (Figure 4B); because of this arrangement, Invasins are categorized in Inverse autotransporter (Ve) family (Leo et al., 2014; Oberhettinger et al., 2012).

InvasinA

Invasin (InvA) is only expressed by the two enteropathogenic species of *Yersinia*, *Y. pseudotuberculosis* and *Y. enterocolitica* (Simonet et al., 1996). The *inv* gene codes for InvA which has a molecular weight of 92 kDa and 106 kDa in *Y. enterocolitica* and *Y. pseudotuberculosis,* respectively (Isberg and Leong, 1990). InvA is expressed at low temperatures, e.g. in stored foods which is beneficial for rapid ingestion of *Yersinia* by the host cells (Grassl et al., 2003).

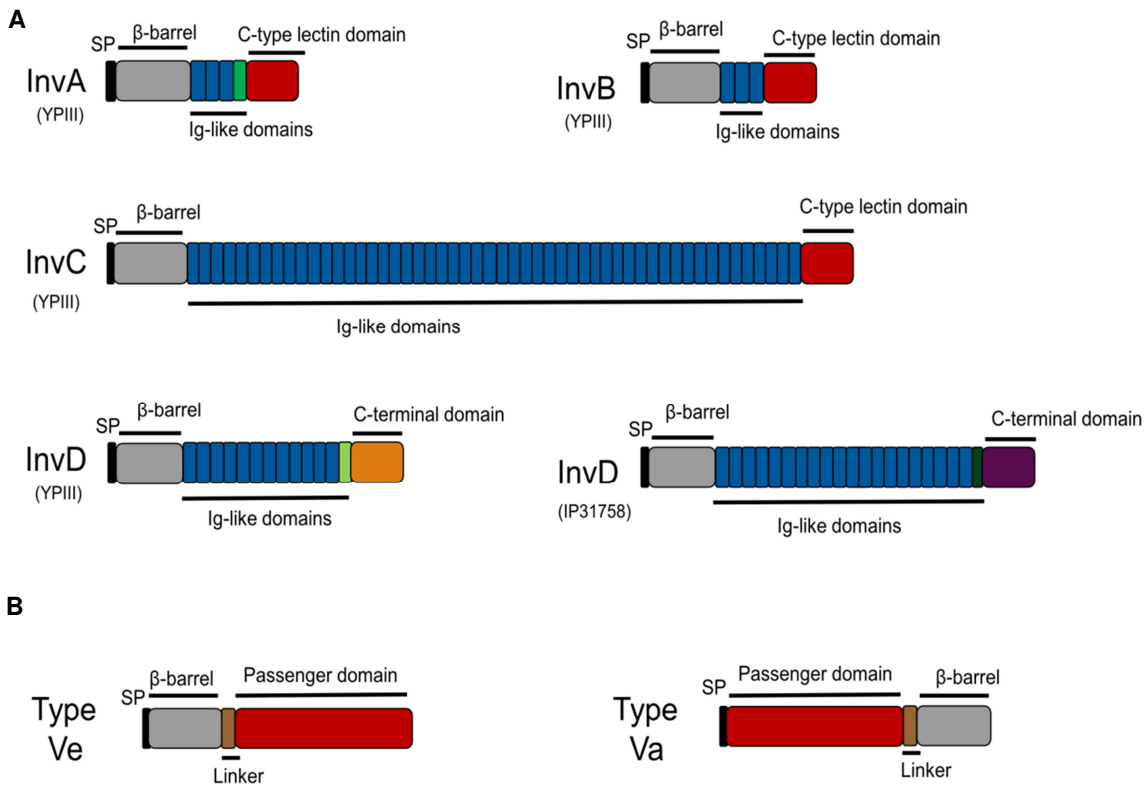

Figure 4: Invasin-like proteins of *Y. pseudotuberculosis.* A) Domain architecture of Invasin-like inverse autotranspoter proteins of YPIII and IP31758 strain of *Y. pseudotuberculosis*. Their N terminus region consists of an export signal peptide (SP; in black) followed by the β-barrel (gray). The part that projects out from the β-barrel consists of repetitive Ig-like domains (blue/green) and a C-terminal domain. B) Comparison of Inverse autotransporters (Type Ve) with classical autotransporters (Type Va). Type Va autotransporters consist of an N-terminal signal peptide which mediates the export of polypeptide into the periplasm, followed by the passenger domain (red), which forms the extracellular domain, the linker (brown) and the β-barrel which is responsible for translocation of the passenger domain. Type Ve contains the same domains but in reversed order.

Y. pseudotuberculosis InvA is responsible for binding to β_1- integrins and causes bacterial uptake. The crystal structure of InvA with four BIg like domains and the adhesion domain has already been solved (Hamburger, 1999) (Figure 5). The crystal structure is composed of four domains (D1-D4), which adopt a fold that is typical of the eukaryotic Immunoglobulin superfamily (IgSF) fold. The D5 domain has a folding topology related to the C-type lectin domain.

Although InvA does not possess the RGD motif, the overall topology and the substrate specificity of the D5 domain are similar to that of fibronectin. Both D4 and D5 together form a superdomain that is essential and constitutes the minimum motif required for integrin binding (Leong et al., 1990). InvA binds to integrins with the affinity that is ~100 times higher than that of fibronectin binding.

The D2 domain of InvA is responsible for homotypic interaction which increases the efficiency of bacterial uptake by multimerization. In contrast, this domain is absent in *Y. enterocolitica;*

hence InvA of *Y. enterocolitica* is unable to form multimers and the invasion process is much less efficient as compared with *Y. pseudotuberculosis* invasion (Dersch and Isberg, 2000).

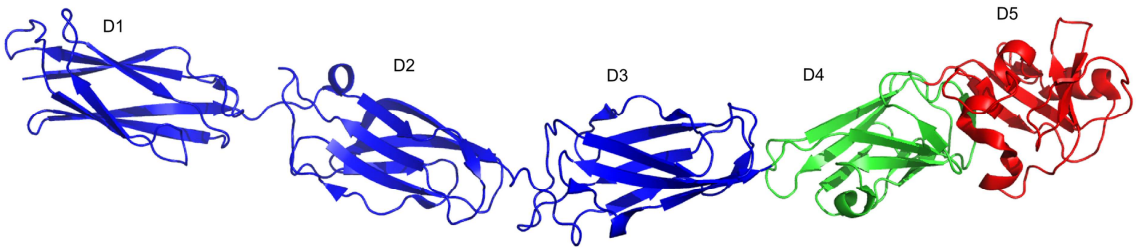

Figure 5: Crystal structure of InvA. Cartoon representation of the crystal structure of InvA (PDB ID:1CWV) consisting of the N-terminal Big domains (D1-D3, shown in blue and D4, shown in green) and the C-terminal C-type lectin domain (D5, shown in red) (Hamburger, 1999).

The expression pattern of InvA is temperature regulated with the maximum expression at lower temperature (20-25°C) and minimal expression at 37°C. Regulation occurs at the transcriptional level by RovA. RovA belongs to MarR/SlyA family of the transcriptional regulators found in *Enterobacteriaceae* (Revell and Miller, 2000). Binding of RovA to the AT rich region of the *inv* promoter at 25°C activates *inv* expression (Heroven et al., 2004). Binding of RovA is temperature dependent and is subject to silencing by the nucleoid-associated H-NS protein. At higher temperatures, RovA gets partially defolded which reduces its DNA binding capability. Transcription of RovA is also repressed by RovM. RovM belongs to the LysR type transcriptional regulator family. It interacts with the DNA region 30 bp upstream of *rovA* promoter, which represses RovA transcription. RovM in turn is controlled by global post transcriptional acting system of *Yersinia* called Carbon Storage Regulator (Csr) system (Heroven and Dersch, 2006; Heroven et al., 2008).

InvasinB

InvasinB (InvB) was recently identified in YPIII. It encodes for 1075 aa protein and consists of an N-terminal β-barrel domain, three Big-like domains and the C-terminal adhesion domain (Figure 4A). InvB is also present in IP31758, IP32953 and PB1/+ strain of *Y. pseudotuberculosis* but is absent in *Y. pestis*. Due to its high sequence similarity with InvA and Intimin, this protein is termed as Ifp (Intimin family protein) in IP32953 strain. Intimin is the adhesin expressed by enteropathogenic *E. coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC). The C-terminal domain of Intimin adopts a C-type lectin like fold and also shares structural similarity with that of InvA. InvB is expressed at 37°C and binds to human, murine and porcine intestinal cells. Furthermore, it plays a role in pathogenesis by promoting

invasion and replication in lymphatic tissues and organs (Pisano et al., 2012; Strong et al., 2011).

InvasinC

A very large protein InvasinC (InvC), comprising of 5337 aa was identified in the YPIII strain of *Y. pseudotuberculosis.* It consists of an N-terminal β-barrel domain, 47 BIg-like domains and the C-terminal adhesion domain (Figure 4A). It is present in IP31758, IP32953 and PB1/+ strain of *Y. pseudotuberculosis,* but the number of BIg-like domains varies significantly in different strains. InvC is highly homologous to *Y. pestis* Ilp (Intimin/Invasin-like protein). Like InvB, InvC was also shown to promote binding to epithelial cells. Deletion of InvC from *Y. pseudotuberculosis* does not significantly affect the host cell binding while deletion of Ilp reduces the dissemination capability of *Y. pestis* (Pisano et al., 2012; Seo et al., 2012).

In addition to InvA, InvB and InvC, two novel Invasins have been recently identified in the strains YPIII and IP31758 of *Y. pseudotuberculosis.* Bioinformatics analysis reveals that the C-terminal domain of these Invasins shares no sequence similarity with the other three Invasins. Apart from that, not much analysis has been done (Pisano et al., 2012).

InvasinD from YPIII strain of *Y. pseudotuberculosis*

InvasinD (InvD) was recently identified in YPIII strain of *Y. pseudotuberculosis.* It is a 1976 aa protein, comprising of an N-terminal β-barrel domain, 13 Ig-like domains and a unique C-terminal adhesion domain (Pisano et al., 2012) (Figure 4A).

InvasinD from IP31758 strain of *Y. pseudotuberculosis*

In addition to InvB, InvC and InvD (YPIII strain), *invD* gene encoding 2795 aa protein has been identified in IP31758 strain of *Y. pseudotuberculosis*. The first 1730 aa of InvD of *Y. pseudotuberculosis* IP31758, corresponding to BIgs, are highly homologous to InvD from *Y. pseudotuberculosis* YPIII, but the C-terminal domain is significantly different (Pisano et al., 2012) (Figure 4A). Due to the significant difference in the C-terminal domain, which in adhesins typically mediates interaction with host molecules, InvD *Y. pseudotuberculosis* IP31758 will hereafter be referred to as (InvasinE) InvE, a novel member of the Invasin-family.

1.5.2 YadA

YadA is a lolli-pop shaped protein present on the cell surface of *Yersinia* that belongs to the family of trimeric autotransporters (TAA) (Hoiczyk et al., 2000). YadA is located on the 70 kb

virulence plasmid of *Yersinia* and is expressed by *Y. enterocolitica* and *Y. pseudotuberculosis* at 37°C immediately after *Yersinia* crosses the intestinal barrier. *yadA* in *Y. pestis* is not expressed due to single nucleotide deletion which leads to a frame shift mutation and a premature stop codon. Inactivation of YadA in *Y. enterocolitica* leads to a virulence in mouse infection model whereas inactivation of YadA in *Y. pseudotuberculosis* does not affect virulence (Rosqvist et al., 1988).

YadA consists of an N-terminal extracellular region called the ''passenger region'' followed by a C-terminal membrane anchor. The passenger region consists of the head domain, followed by the neck and a rigid coiled-coil pillar like stalk domain. The structure of the head along with the neck region of YadA from *Y. enterocolitica* ($YadA_{ent}$) has been solved (Nummelin et al., 2004). The head domain of $YadA_{ent}$ (aa 24-196) is solely composed of β-sheets that adopt a left-handed parallel β roll (LPBR) arrangement. The LPBR is composed of a 13-16 residue repeat motif, containing an NSVAI**G**XXS repeat with a highly conserved G. NSVAIGXXS motif forms the hydrophobic interior of the trimer. The neck region connects the LPBR region to the coiled coil stalk domain (Figure 6A) (Nummelin et al., 2004).

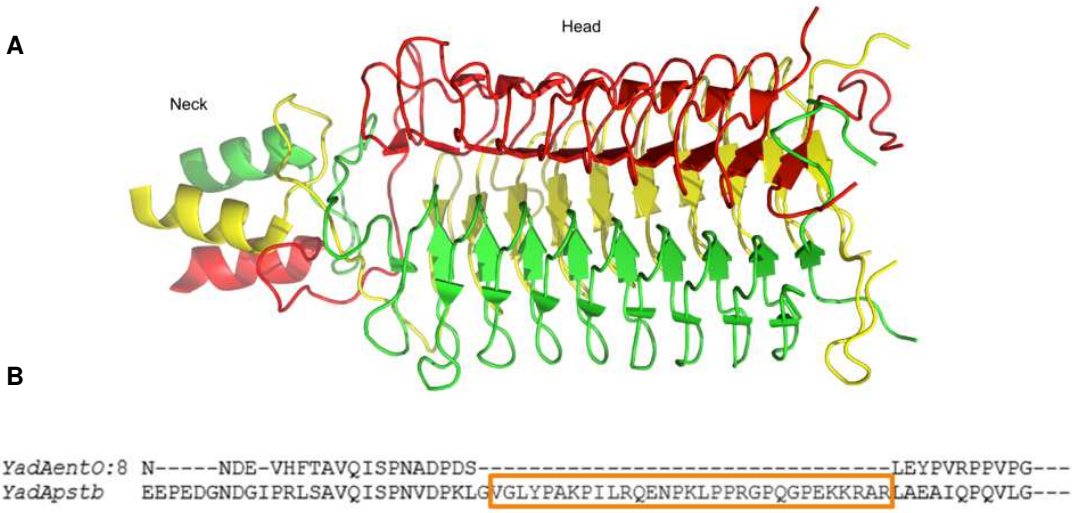

Figure 6: Structure of *Y. enterocolitica* YadA. A) Cartoon model of the trimeric $YadA_{ent}$ crystal structure (PDB ID:1P9H) consisting of the N-terminal head and the C-terminal neck region (Nummelin et al., 2004). B) Alignment of the N-terminal section of the head domain of YadA of *Y. enterocolitica* (serotype O:8) and *Y. pseudotuberculosis*. 31 aa insertion that is unique to $YadA_{pstb}$ is marked in an orange box.

YadA is a multifunctional protein. In *Y. enterocolitica*, YadA mediates binding to extracellular matrix (ECM) components such as laminin, collagen and fibronectin. It preferably binds to collagen I, II, III and V. $YadA_{ent}$-collagen interaction is quite stable and can tolerate extremes of pH, temperature and is resistant to proteases and chaotropic agents (Emody et al., 1989). Binding affinity of $YadA_{ent}$ to other ECM components, such as fibronectin and laminin is low

as compared to collagen. $YadA_{ent}$ does not bind to plasma fibronectin but only to cellular fibronectin (Emody et al., 1989; Heise and Dersch, 2006).

In contrast, YadA of *Y. pseudotuberculosis* ($YadA_{pstb}$) preferentially binds fibronectin and mediates cell invasion. This is due to the presence of an additional 31 aa (amino acids 53-83) at the N-terminal region of the head domain of $YadA_{pstb}$ (Figure 6B). Deletion of this region results in the loss of cell invasion property of *Y. pseudotuberculosis* and bacteria gains high affinity for collagen and laminin binding and hence this region has been named as "uptake domain" but the structural basis for this is unclear (Heise and Dersch, 2006). In contrast to InvA, YadA- β_1 integrins interaction is not direct and occurs via an ECM bridging mechanism. YadA interaction with the ECM molecules which are present on the epithelial cells, triggers signaling cascades via the β_1- integrin receptors bound to the ECM molecules (Eitel and Dersch, 2002).

Besides its function as an adhesin, $YadA_{ent}$ also provides protection to the bacteria by blocking the innate immune system. It provides serum resistance to *Yersinia* by binding to complement factors like factor H, C3b and C4BP thereby preventing complement mediated lysis (Biedzka-Sarek et al., 2008a; Kolodziejek et al., 2010).

1.5.3 Ail

Ail is a 17 kDa outer membrane protein, chromosomally located and present in all three pathogenic *Yersinia* species. It is responsible for host cell adhesion and evading host immune response. Ail is expressed at 37°C during the stationary phase (Pierson and Falkow, 1993). The crystal structure of Ail from *Y. pestis* has been solved (Yamashita et al., 2012) (Figure 7).

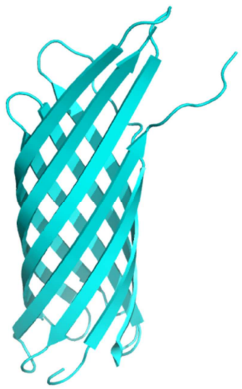

Figure 7: **Crystal structure of Ail.** Structure of Ail (PDB ID:3QRA) shown in cartoon model (Yamashita et al., 2012).

It forms an eight stranded anti-parallel β-barrel structure that spans the outer membrane with four extracellular loops that are responsible for cell adhesion. It is structurally similar to the *E. coli* invasion protein OmpX (Yamashita et al., 2012) (Figure 7).

In *Y. enterocolitica* and *Y. pestis,* Ail mediates attachment to ECM proteins such as laminin, fibronectin and heparin sulfate, while *Y. pseudotuberculosis* Ail does not mediate binding to

ECM (Miller et al., 2001). Ail is responsible for binding C4BP and factor H, thus blocks the complement activation pathway which leads to bacterial survival. Unlike YadA, it can be masked by the lipopolysaccharide O antigen in *Y. pseudotuberculosis* and *Y. enterocolitica* due to the small size of the extracellular regions of Ail. Thus the serum resistance, adhesion and internalization activity of Ail is dependent on the LPS core structure (Biedzka-Sarek et al., 2008; Kolodziejek et al., 2010).

1.5.4 Pla

Plasminogen activator (Pla) is the outer membrane protein of *Y. pestis* that belongs to the family of OM proteases/adhesins. It is encoded by pCP1 plasmid of *Y. pestis*. Deletion of this gene reduces the virulence of *Y. pestis* in supporting pneumonic plague and bubonic plague.

The structure of Pla (312 aa) forms a narrow β-barrel and is composed of ten antiparallel β-strands. The strands are connected by short turns at the periplasmic side (bottom) and by loops at the extracellular side (top) (Eren et al., 2010) (Figure 8). Pla is actively associated with virulence as it cleaves plasminogen to plasmin and in turn activates matrix metalloproteinases which leads to degradation of networks in extracellular matrices.

Figure 8: Crystal structure of Pla. Structure of Pla (PDB ID: 2X55) shown in cartoon model (Eren et al., 2010).

This activity helps in dissemination of bacteria from the primary infection site by preventing the formation of fibrin around them and also in inhibiting the formation of microabscesses. Pla has also been shown to cleave complement protein C3, providing resistance to bacteria against complement mediated lysis. However, *Y. pestis* strain that lacks Pla is resistant to this complement system indicating that this is not its major function (Sodeinde et al., 1992). At 37°C, Pla could mediate translocation of Yop proteins to host cells by positioning the T3SS in such a manner that Yop proteins could be injected into the host cell (Felek et al., 2010).

1.5.5 PsaA

PsaA (pH six antigen A) is encoded by the chromosomal region of *Yersinia* and contributes to its virulence. Several subunits of 15 kDa PsaA form a fibrillary structure called pH6 antigen.

The structure of PsaA adopts an Ig-like fold and consists of seven β-strands (Bao et al., 2013) (Figure 9). It is produced inside the macrophages in an acidic intracellular environment and extracellular environment in abscesses such as buboes (Lindler and Tall, 1993; Lindler et al., 1990). pH6 of *Y. pestis* and *Y. pseudotuberculosis* promotes hemagglutination and also promotes thermoinducible binding to tissue culture cells (Yang et al., 1996).

Figure 9: Crystal structure of PsaA. Structure of PsaA (PDB ID: 4F8O) shown in cartoon model (Bao et al., 2013).

pH6 of *Y. pestis* binds to lipoproteins which prevents recognition of the bacteria by the host immune system (Makoveichuk et al., 2003). *Y. enterocolitica* possesses Myf, a homologue of pH6 which also forms a fibrillary structure. The exact function of Myf is however not clear (Yang and Isberg, 1997).

1.6 Bacterial secretion system

Gram-negative bacteria possess a variety of secretion machines to export wide ranges of substrates, including DNA, proteins, and small molecules across the bacterial cell envelope.

In Gram-negative bacteria, these nano-machines are of six different types clustered in two subgroups (Figure 10)-

i) Secretion machines that span the double membrane- Type 1 Secretion System (T1SS), T2SS, T3SS, T4SS and T6SS.
ii) Secretion machine that spans only the outer membrane (OM) of the bacteria- T5SS (Abdallah et al., 2007).

1.6.1 T1SS

The T1S system delivers a large variety of molecules from the cytoplasm to the extracellular environment in a single step, independent of the SecYEG translocon. This system is composed of three components: i) IMC (Inner Membrane Component) that belongs to the ATP binding cassette transporter, ii) a periplasmic adaptor protein and iii) TolC, which forms

the outer membrane channel. Binding of substrates to the IMC triggers the sequential assembly of IMC, periplasmic adaptor and TolC. This leads to the opening of TolC and the substrate is released into the extracellular space (Costa et al., 2015).

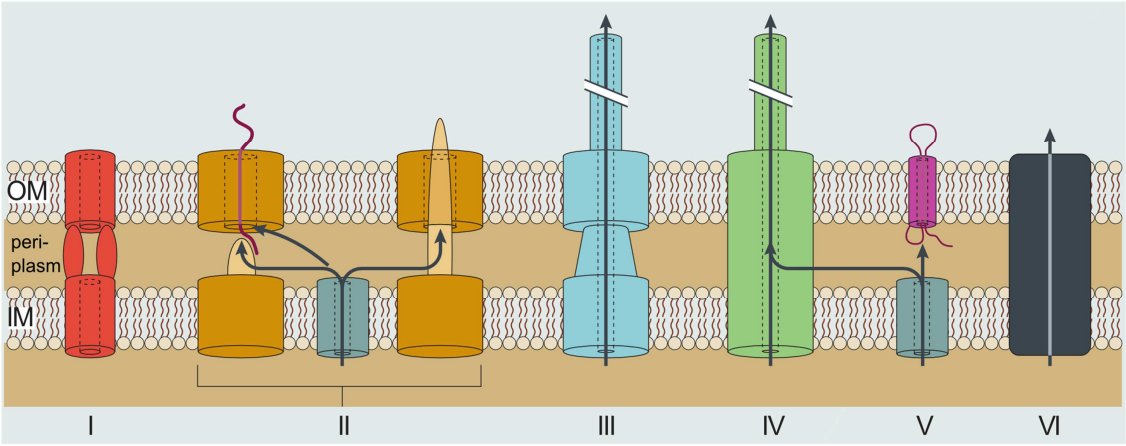

Figure 10: Bacterial secretion systems.

Type 1 secretion system is composed of the inner membrane (IM) component, membrane fusion protein and the outer membrane (OM) component. Substrates are transported into the extracellular space by a one-step mechanism.

Type 2 secretion system spans the double membrane of the bacteria. Substrate is transported to the periplasm by the Sec translocon machinery. Proteins fold in the periplasm and are transported to the extracellular space.

Type 3 secretion system forms a channel that spans the double membrane and extends to contact host cell's plasma membrane.

Type 4 secretion system also transports substrate directly to the host cell by a one or a two-step mechanism.

In Type 5 secretion system, the SEC translocon is used to transfer substrates to the periplasm. A β-barrel domain inserts into the OM and the passenger domain is translocated through the β-barrel.

Type 6 secretion system spans the cell envelope of bacteria. The mechanism of secretion resembles the contraction mechanism of phage tails. (Adapted from (Abdallah et al., 2007)).

1.6.2 T2SS

The T2S system is composed of 12-15 components that transport a variety of proteins (such as toxins, hydrolyzing enzymes that are mainly involved for bacterial survival in their niche) to the extracellular environment in two steps. In the first step, unfolded polypeptides are transported to the periplasm by the SecYEG translocon. Folded proteins are then transported from the periplasm to the extracellular space in a separate step which is mediated by the secreton. The secreton is formed by the OM component (secretin) and the IM component that pushes the substrate through the OM pore (Costa et al., 2015; Sandkvist, 2001). All pathogenic and non-pathogenic species of *Yersinia* possess one or two copies of T2SS, but not much is known about the role of this system in *Yersinia* (von Tils et al., 2012).

1.6.3 T3SS

T3SS, also called injectisome, is found in a variety of pathogenic Gram-negative bacteria. T3SS is structurally and functionally analogous to bacterial flagella (Blocker et al., 2003). It

transfers bacterial effector proteins directly into the host cell cytoplasm, bypassing the extracellular environment in a Sec-independent manner. Within the host cell, these effector proteins modulate or interfere with the host cell processes and promote bacterial invasion and colonization. The system is made up of 20 different bacterial components that form a syringe like structure which spans the inner membrane, periplasm and the outer membrane of bacteria, with the needle extending out of the bacteria. Secretion of bacterial proteins occurs, when the tip of the complex contacts the host cell plasma membrane. This contact triggers the release of translocator proteins that form pores in the host cell plasma membrane. The needle tip attaches to the pore, thus completing a channel through which effector proteins are passed from the cytoplasm into the host cell (Galán et al., 2014). All pathogenic species of *Yersinia* use this system to deliver effector proteins into the host cell, thereby they gain resistance to phagocytosis and escape the host immune system (Cornelis, 2002).

1.6.4 T4SS

T4SS is composed of twelve different proteins that together form the entire double membrane spanning secretion apparatus. T4SS is classified into three subfamilies depending on its function. The first subfamily delivers DNA to other bacteria, fungi or plant cells by cell to cell contact. The second subfamily involves DNA uptake from and release into the extracellular environment and promotes genetic exchange. The third system is the effector translocator system that contributes to transfer of effector proteins to eukaryotic cells (Cascales and Christie, 2003).

1.6.5 T5SS

T5SS also called as the autotransporter system, is a single membrane spanning secretion system. This machine requires the SecYEG translocon to mediate export of proteins to the periplasm. Proteins that transport through this mechanism consist of a β-barrel domain also called translocation domain, that inserts into the OM and the passenger domain is secreted through the pore of the β-barrel. This system is divided into five subclasses-Va, Vb, Vc, Vd and Ve (Leo et al., 2012).

Va secretion system- It uses a classical form of autotransport in which the protein is transported to the periplasm by the Sec machinery via its N-terminal signal peptide. In the periplasm, various chaperones such as SurA, Skp, FkpA and DegP are involved in quality control of the transported protein. The signal peptide is cleaved off and the C-terminal β-barrel domain inserts into the OM with the help of Bam complex. The linker region then forms a hairpin inside the pore of the barrel which enables the transport of the passenger domain into the extracellular space through the pore of the β-barrel (Leo et al., 2012).

Vb secretion system- This system is also known as the Two-Partner Secretion System (TPSS). As the name suggests, the passenger domain (referred to as TpsA) and the translocation domain (TpsB) are present as separate polypeptide chains which are transported to the periplasm by the Sec machinery. In the OM, TpsB protein folds into a β-barrel structure. TpsA protein contains the two-partner secretion (TPS) domain that targets the protein to TpsB, TpsA protein is then translocated through the OM into the extracellular space (Henderson et al., 2004; Leo et al., 2012).

Vc secretion system- This secretion system is also called trimeric autotransporters, the prototype of which is YadA. As the name suggests, these are homotrimeric proteins, with the translocation domain and the passenger domain being made up of three polypeptide chains. The system follows a similar transport mechanism like the Va system, but the major difference is the presence of three instead of one polypeptide chain (Leo et al., 2012).

Vd secretion system-This system is similar to that of Vb secretion system, the major difference being that the two polypeptides are fused in a single polypeptide (Leo et al., 2012).

Ve secretion system–This system is known as the inverse autotransporter family, the prototype of which is the Invasins. In contrast to Va autotransporters, their topology is inverted. Their translocation domain is at the N-terminus while the C-terminus contains the passenger domain (Figure 4B) (Henderson et al., 2004; Leo et al., 2014).

1.6.6 T6SS

T6SS can secrete proteins into both prokaryotic and eukaryotic cells (Gueguen et al., 2013). The role of T6SS has been reported in virulence, host immunomodulation, stress-sensing and biofilm formation (Weber et al., 2009). T6SS was earlier considered to play a role in virulence but recent research also points out the role of T6SS in interbacterial interactions (Schwarz et al., 2010).

Bacterial T6SS is structurally analogous to a bacteriophage (Ho et al., 2014). Although the overall architecture of T6SS is well conserved, this system has different functions depending on the need of individual species.

Most of the Gram-negative bacteria have one or two copies of the T6SS cluster in their genome, but *Y. pseudotuberculosis* and *Y. pestis* have 4-5 clusters, suggesting multiple functions or specificities of T6SS in different hosts (Bingle et al., 2008). In *Y. pseudotuberculosis*, there are four complete and two incomplete loci, while in *Y. enterocolitica* only one copy is present (Zhang et al., 2011). To respond to various environmental changes, T6SS clusters of *Yersinia* are regulated at transcriptional, post-transcriptional and translational levels by a variety of proteins. T6SS4 is under control of one single promoter, while T6SS1, T6SS2 and T6SS3 clusters are divided into two to three putative operons (Zhang et al., 2011). T6SS4 promotes survival of *Yersinia* in high osmolarity

conditions and also provides resistance to deoxycholate. This system in turn is regulated by the response regulator OmpR which binds and activates the promoter region of T6SS4 (Gueguen et al., 2013). T6SS expression is temperature dependent. T6SS2, T6SS3 and T6SS4 of *Yersinia* are strongly induced at moderate temperature in contrast to T6SS1 loci which is expressed at 37°C (Zhang et al., 2011).

1.7 Csr system of *Yersinia*

To achieve efficient infection, the virulence factors must be expressed at the correct condition and at the correct time. These changes are governed by complex regulatory networks in *Yersinia* that control the expression of these factors depending on the condition. This is achieved by regulation at transcriptional, post-transcriptional and post-translational levels by a variety of regulatory systems.

One such system is the Csr (Carbon storage regulator) system in *Yersinia*. This is a post-transcriptional regulatory system of *Yersinia* that is composed of the RNA binding protein referred to as CsrA, and two non-coding RNAs CsrB and CsrC (Figure 11).

CsrA is a homodimer that regulates the translation of its target mRNAs by blocking the ribosome binding site at or near the Shine-Dalgarno (SD) sequence, and thus competes with binding of the 30S ribosomal subunit, thereby blocking translation. CsrA activity is controlled by two non-coding RNAs CsrB and CsrC that can bind multiple CsrA molecules, thus controlling the availability of CsrA in the cell and preventing binding of CsrA to its target mRNA.

Csr is a part of the regulatory RovM-RovA-InvasinA cascade. CsrA positively regulates RovM by a not clearly known mechanism, thus blocking the expression of RovA, which in turn blocks the expression of InvA (Heroven et al., 2008, 2012) (Figure 11).

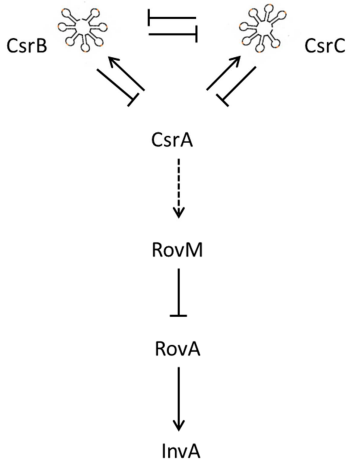

Figure 11: Csr system of *Yersinia*. Two non-coding RNAs CsrB and CsrC antagonize CsrA function and also counter-regulate each other. Up-regulation of CsrA increases the expression level of RovM, leading to RovA repression and hence InvA synthesis is sequestered. Dashed lines indicate indirect interaction while solid lines indicate direct interaction (modified from (dissertation Stephanie Christine Seekircher, 2014)).

1.8 RovC

A novel transcriptional regulator has been identified in Prof. Petra Dersch's group (HZI) that represses the transcription of CsrC and thus designated as regulator of virulence associated with CsrC (RovC). This protein is unique to *Yersinia*. It is highly conserved in *Y. pestis* and in *Y. pseudotuberculosis* and shares no sequence homology to any other protein in the database and also has no conserved DNA binding domain. RovC sequesters CsrC transcription, which leads to increased CsrA levels, thus positively affecting RovM expression which in turn represses RovA and hence InvA. RovC is thermoregulated and shows maximal expression at 25°C in stationary growth phase. It requires factors for its expression that are specific for *Yersinia*. RovC is located directly downstream of T6SS4 gene, encoded in opposite direction and activates T6SS4 transcription (dissertation Stephanie Christine Seekircher, 2014) (Figure 12).

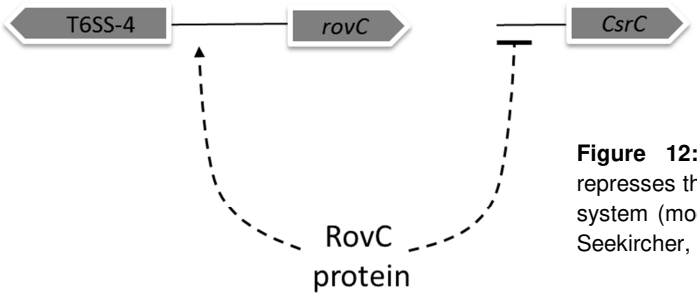

Figure 12: RovC regulation in *Yersinia*. RovC represses the transcription of CsrC but activates T6SS4 system (modified from dissertation Stephanie Christine Seekircher, 2014).

2 Aims

Yersinia pseudotuberculosis armour themselves with a variety of virulence factors that enable them to adhere to host cells and colonize host tissues for their survival. Their expression levels in *Yersinia* are regulated by several regulatory proteins in response to the rapidly changing environment. Thus, the aims of the project were to structurally and functionally characterize adhesins and the virulence regulator of *Y. pseudotuberculosis.*

As outlined above, InvA is a major virulence factor of enteropathogenic *Yersinia* required for adhesion to and invasion of epithelial cells. InvA has been extensively characterized at both structural and functional levels. The first objective of this work was to structurally and biochemically characterize the two recently identified Invasins, InvD and InvE, that share no sequence similarity with other known Invasins. Since Invasins share a common architecture, the molecular details of the adhesion domain are needed to provide the functional information. Thus, the aim was to gain atomic and molecular insights into the structure of InvD and InvE by X-ray crystallography and to identify their host-interaction partner by various biochemical and molecular biology experiments. The identified interactions should be analyzed *in vitro,* and if confirmed, InvD/InvE should be co-crystallized with their respective targets to characterize the interaction surface. These studies will increase our knowledge of the diverse Invasins of *Yersinia*, and will contribute to the understanding of the role of Invasins in their host system during infection.

The second aim of this work was to provide the molecular mechanism for the differences in the function of YadA present in *Y. pseudotuberculosis* and *Y. enterocolitica*. For that purpose, the structure of the head domain of $YadA_{pstb}$ should be solved by X-ray crystallography and compared to known YadA structure from *Y. enterocolitica*. The differences in two structures should be analyzed in the context of their function.

The third goal of this work was to structurally and biochemically characterize the virulence regulator RovC in the context of DNA binding. For that purpose, RovC should be produced in high purity and quantity for crystallization experiments. Then, the structure of RovC alone and in complex with DNA should be solved to understand its DNA binding mechanism. Additionally, biochemical and biophysical methods should be used to characterize RovC and its DNA binding activity.

3 Materials and methods

3.1 Materials

3.1.1 Buffers and solutions

Table 1: Buffers and solutions.

Ni-NTA Lysis buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl, 1 mM Imidazole pH 8, 5 mM β-mercaptoethanol (βME), 5 mM $MgCl_2$, 1 mM PMSF, DNAse (40 µl at 1 mg/ml per litre of culture)
Ni-NTA wash buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl, 5 mM Imidazole pH 8, 5 mM βME
Ni-NTA elution buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl, 250 mM Imidazole pH 8, 5 mM βME
Streptactin Lysis buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl, 5 mM βME, 5 mM $MgCl_2$, 1 mM PMSF, DNAse (40 µl at 1 mg/ml per litre of culture)
Streptactin wash buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl, 5 mM βME
Streptactin elution buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl,

	5 mM βME, 3 mM D-Desthiobiotin
TEV protease cleavage buffer	50 mM TRIS pH 8 200 mM NaCl, 5 mM βME
Gel filtration buffer	20 mM HEPES pH 7.4, 200 mM NaCl, 5 mM DTT
Lysine methylation dialysis buffer	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 5 mM βME
Lysine methylation gel filtration buffer	20 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5 mM βME
Dialysis buffer	50 mM TRIS-HCl pH 8, 200 mM NaCl, 5 mM βME
PBS buffer (10x)	1.4 M NaCl, 30 mM KCl, 120 mM Na_2HPO_4, 20 mM KH_2PO_4, pH 7.4
PBS-T	1x PBS, 0.1% (v/v) Tween-20
MPBST	PBST, 2% skim milk
TBS (10x)	100 mM TRIS-HCl pH 7.5, 1.5 M NaCl
TBS-T	1x TBS, 0.1% (v/v) Tween-20
Transfer buffer	25 mM TRIS-Base pH 8.0,

	192 mM Glycine, 10% Methanol (v/v)
Blocking buffer	5% (w/v) Skim milk in TBS-T
Alkaline Phosphatase (AP) buffer	100 mM TRIS-HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$
Alkaline Phosphatase staining solution	10 ml AP buffer, 66 µl NBT substrate (stock solution-50 mg/ml), 33 µl BCIP substrate (stock solution-50 mg/ml)
Coupling buffer	0.2 M $NaHCO_3$ pH 8.6, 0.5 M NaCl
TAE buffer (50x)	2 M TRIS-Base, 1 M Acetic acid, 50 mM EDTA
DNA loading buffer (10x)	10 mM TRIS-HCl pH 7.5, 0.05%(w/v) Bromphenol blue, 1 mM EDTA, 5% Glycerol
SDS loading buffer (2x)	100 mM TRIS-HCl pH 6.8, 4% (w/v) SDS, 20% (v/v) Glycerol, 200 mM DTT, 0.2% (w/v) Bromophenol blue
SDS-PAGE running buffer	25 mM TRIS-Base pH 8.5, 192 mM Glycine, 0.1% (w/v) SDS
Coomassie blue staining solution	0.1% (w/v) Coomassie brilliant blue R-250, 40% (v/v) Ethanol, 10% (v/v) Acetic acid
Destaining solution	40% (v/v) Ethanol 10% (v/v) Acetic acid

3.1.2 Culture media and antibiotics

Table 2: Media and antibiotics used in this work.

Luria Bertani (LB) medium	10 g/L Bacto-tryptone, 5 g/L Yeast extract, 7 g/L NaCl
Terrific Broth (TB) medium	12 g/L Bacto-tryptone, 24 g/L Yeast extract, 4 ml/L Glycerol, 12.54 g/L K_2HPO_4, 2.31 g/L KH_2PO_4
double Yeast Trypton Broth (dYT) medium	16 g/L Bacto-tryptone, 10 g/L Yeast extract, 5 g/L NaCl
LB plate-Agar	10 g/L Bacto-tryptone, 5 g/L Yeast extract, 7 g/L NaCl, 16 g/L Bacto-Agar
SOC medium	20 g/L Bacto-tryptone, 5 g/L Yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ 20 mM Glucose
dYT-GA medium	dYT medium, 100 mM Glucose, 100 µg/ml Ampicillin
Minimal medium	1 g/900 ml NH_4Cl, 3 g/900 ml KH_2PO_4, 4 g/900 ml Na_2HPO_4
Additional solution	22 g/100 ml Glucose,

	0.61 g/100 ml $MgSO_4$, 11.2 mg/100 ml Thiamine-HCl, 10.4 mg/100ml $Fe_2(SO_4)_3.7H_2O$
Se-Met Medium	Minimal Medium + Additional solution
RPMI 1640 medium	Ordered from Gibco Life technologies, Contains 1% Glutamax-1
HEp-2 medium	RPMI 1640 (supplemented with 1% Glutamax-1), 10% Newborn Calf Serum (NCS)
Antibiotics	
Kanamycin (1000x)	30 µg/ml in deionized water
Chloramphenicol (1000x)	34 µg/ml in 70% Ethanol (v/v)
Ampicillin (1000x)	30 µg/ml in deionized water

3.1.3 Enzymes

All enzymes listed in Table 3 were used according to manufacturer´s instructions.

Table 3: Enzymes used in this work.

KOD Hot Start DNA Polymerase	Novagen
T4 DNA ligase	New England Biolabs
Restriction Enzymes	New England Biolabs
Calf Intestinal Phosphatase	New England Biolabs
DNaseI	Roche
T4 Polynucleotide Kinase	Thermo Fischer Scientific Inc.
2x Red PCR Master Mix	pjk GmbH

3.1.4 Bacterial strains

Table 4: *E. coli* strains used in this work.

Name	**Description**	**Source**
DH5α	F-Φ80*lacZ*ΔM15 Δ(*lacZYA-arg*F) U169 *recA1endA1hsd*R17 (rk-, mk+) *phoAsup*E44λ -*thi*-1*gyr*A96 relA1	Invitrogen

Rosetta2 DE3	F- *ompThsd*S_B(r_B- m_B-) *gal dcm* (DE3) pRARE2 (Cam^R)	Invitrogen

3.1.5 Antibodies

Antibodies used in western blot and enzyme linked immunosorbent assays.

Table 5: Antibodies used in this work.

His-AP conjugate	Amersham Bioscience
Anti-rabbit HRP coupled	Jackson Immuno Research
Anti-flag antibody produced in rabbit	Sigma Aldrich
9E10 mouse anti-myc antibody	Yumab GmbH, Braunschweig
Goat anti-mouse IgG-HRP coupled antibody	Sigma Aldrich

Antibodies used for flow cytometry experiments

Fluorochrome-conjugated anti-CD3 (17A2), anti-CD19 (6D5), anti-CD11b (M1/70), anti-CD11c (N418), anti-CD49b (DX5), anti-γδTCR (eBioGL3) and anti-flag (L5, DYKDDDDK tag epitope) were purchased from eBioscience (San Diego, CA, USA) and Biolegend (San Diego, CA, USA).

3.1.6 Protein and nucleotide standards

Applied size standards are indicated in Table 6.

Table 6: Molecular size standards.

Gene Ruler DNA ladder mix	Thermo Scientific
PageRuler unstained protein ladder	Fermentas
PageRuler Plus prestained protein ladder	Fermentas

3.1.7 Columns and resins

Table 7: Resins used in this work.

Ni Sepharose® 6 Fast Flow	GE Healthcare
Strep-Tactin® Sepharose® 50% suspension	Iba
Strep-Tactin® Macroprep® 50% suspension	Iba
Anti-Flag M2 magnetic beads	Sigma Aldrich

Table 8: Columns used in this work.

HiLoad 16/600 Superdex 200 prep grade	GE Healthcare
HiLoad 16/600 Superdex 75 prep grade	GE Healthcare
Superdex 200 10/300	GE Healthcare
Superdex 75 10/300	GE Healthcare

3.1.8 Kits

All kits listed in Table 9 were used according to manufacturer's instructions.

Table 9: Commercial kits used in this work.

QIAquick PCR purification kit	Qiagen
QIAquick Gel extraction kit	Qiagen
QIAprep Spin Miniprep kit	Qiagen
Proti-Ace	Hampton Research
Proti-Ace-II	Hampton Research
Lanthanide Phasing kit	Jena Bioscience

3.1.9 Crystallization screens

Table 10: List of crystallization screens used.

Additive screen	Hampton Research
$AmSO_4$	Qiagen
Anion	Qiagen
Cation	Qiagen
Cryo	Qiagen
Index	Hampton Research
JCSG CoreI	Qiagen
JCSG CoreII	Qiagen
JCSG CoreIII	Qiagen
JCSG CoreIV	Qiagen
JCSG+	Qiagen
Midas	Molecular Dimensions
Morpheus	Molecular Dimensions
Nucleix screen	Qiagen

PEGs	Qiagen
Protein Complex	Qiagen
Silver bullet	Hampton Research

3.1.10 Oligonucleotides

List of oligonucleotides used in this work to amplify gene of interest are listed in Table 11. Oligonucleotides used to insert mutations are listed in Table 12. Oligonucleotides that were used for sequencing of constructs or plasmids are listed in Table 13. Table 14 depicts the oligonucleotides used to modify plasmids. All the primers were purchased from Eurofins Genomics.

Table 11: Oligonucleotides used for DNA amplification.

Oligo name	**Sequence 5´→3´**	**Restriction site**
InvD		
InvD_G1640_notI_f	aagaatgcggccgcggcaacctgagcaccacgaac	NotI
InvD_N1976_pstI_r	ttttctgcagttagttagatccgg	PstI
InvD_P1738_notI_f	aagaatgcggccgcggtgcattagatgctgcgcgc	NotI
Ig_G1839_pstI_r	ttttctgcagttaccccggcgttattgtcaccatc	PstI
InvA		
InvA_P500_notI_f	aagaatgcggccgccctcagttgacattaacggcggcc	NotI
InvA_I985_pstI_r	ttttctgcagttatattgacagcgcacagagcgg	PstI
InvE		
InvE_G2448_notI_f	aagaatgcggccgcggcaacctgagcaccacgaac	NotI
InvE_A2547_notI_f	aagaatgcggccgcgctcctgttgatctaacggtattaac	NotI
InvE_L2795_kpnI_r	ggggtaccttatagagaacagcttacattgtagtttc	KpnI
InvE_G2536_notI_f	aagaatgcggccgcggcacaacaccagcaacgattaatg	NotI
InvE_P2539_notI_f	aagaatgcggccgccagcaacgattaatgtcattcctg	NotI
InvE_V2544_notI_f	aagaatgcggccgcgtcattcctgctcctgttgatctaac	NotI

RovC

RovC_M1_notI_f	aagaatgcggccgcatgagaaagaagctatataatgacttc	NotI
RovC_D40_notI_f	aagaatgcggccgcgattctgaattaatccaatcagag	NotI
RovC_S47_notI_f	aagaatgcggccgctcagagctggatttgaatgcag	NotI
RovC_S80_notI_f	aagaatgcggccgcagtaaccgttcagttagagttaaactc	NotI
RovC_M103_notI_f	aagaatgcggccgcatgtccaacctaccaggagtcaaac	NotI
RovC_L247_kpnI_r	gggtaccttagaggaagttcaggtagccg	NotI

YadA

YadA_E26_notI_f	aagaatgcggccgcgaggagcccgaggatggcaac	NotI
YadA_E253_kpn_r	gggtaccttattcagccatttctttctttaattg	kpnI

Table 12: Oligonucleotides used to insert mutations/deletions in the gene of interest.

Oligo name	**Sequence 5´→3´**
InvE	
InvE_E2695A/K2696A_f	Gcagcggcgccaggcacaataggcaccaatgttgaaaac
InvE_ E2695A/K2696A _r	Gcctggcgccgctgctgccaagtatgggcgtttgaac
InvE_E2648A_f	Aaagcggcgagcggtgatatggaaaatactaatgtcacc
InvE_E2648_r	Accgctcgccgctttagccctaagggttgtttgcac
InvE_K2646A_f	Agggctgcggcggaaagcggtgatatggaaaatactaatg
InvE_ K2646A _r	Ttccgccgcagccctaagggttgtttgcacaccaatacc
InvE_E2505A/Q2506A_f	Gtgaccgcggcggccagtggcctctataccgcgacc
InvE_E2505A/Q2506A _r	Ggccgccgcggtcacattgtccagcgtccctagagtg
RovC	
RovC_K181A/E182A/E183A_f	Tgcgcggcggcgcaatatttaggtctcttgaagaccatag
RovC_ K181A/E182A/E183A _r	Ttgcgccgccgcgcattccacctctattttgtgattaac
RovC_K233A/K234A _f	Cgtatagcggcggcgaatgcacttattaattacggctacctg
RovC_ K233A/K234A_r	Attcgccgccgctatacgatagcggatttttgctctgac
RovC_K56E_f	Gaaaaagaatggggagtaatgaaatatattgatccttataactctg
RovC_K56E_r	Tactccccattcttttctgcattcaaatccagctctgattg

RovC_R225E_f	tgggtcgaagcaaaaatccgctatcgtataaagaaggcg
RovC_R225E_r	ttttgcttcgacccaactgtcagcagaccattcatttttac
RovC_R229E_f	aaaatcgaatatcgtataaagaaggcgaatgcacttattaattac
RovC_R229E_r	tatacgatattcgatttttgctctgacccaactgtcagc
RovC_I150P_f	ctctatccttatatccctttgaaccttgaatctgatgtttttg
RovC_I150P_r	agggatataggatagaggtttgaatcatcttttaatgcatctgc
RovC_I150P/Y151P_f	ctctatccttatatccctttgaaccttgaatctgatgtttttg
RovC_I150P/Y151P_r	agggataggaggatagaggtttgaatcatcttttaatgcatctgc
RovC_A237E_f	aaggcgaatgaacttattaattacggctacctgaacttcctc
RovC_A237E_r	attaataagttcattcgccttctttatacgatAgcggattttcg
RovC_G242E_kpn_r	ggggtaccttagaggaagttcaggtattcg
RovC_S219E/A220E_f	gaatgggaagaagacagttgggtcagagcaaaaatccgc
RovC_S219E/A220E_r	actgtcttcttccattcatttttaccagttccttacC

Table 13: Oligonucleotides used for sequencing.

Oligo name	**Sequence 5´→3´**
DuetUP1	ggatctcgacgctctccct
DuetDOWN1	gattatgcggccgtgtacaa
MHLacZ-Pro_f	ggctcgtatgttgtgtgg

Table 14: Oligos used to modify plasmids (to insert tags).

Oligo name	**Sequence 5´→3´**
Inserting 3xflag tag at the N terminus in pVP009	**Modified vector is 6xHis_TEV_3xflag_MCS1**
3xFlag_notI_2f	aaagatgacgatgacaagggcggccgcctgcagggtac
3xFlag_notI_4r	cgtcatggtctttgtagtcgccac
3xFlag_notI_1f	ggccgtggcgactacaaagaccatgacggtgattataaagatcatgatatcgattac
3xFlag_notI_3r	cctgcaggcggccgcccttgtcatcgtcatctttgtaatcgatatcatgatctttataatcac

Inserting SUMO tag at the N terminus in pVP009	**Modified vector is 6xHis_SUMO_TEV_MCS1**
SUMO_TEV_notI_r	Attcttgcggccgcctccctggaagtagaggttctcgtcgacaccaatctgttctctgtg
SUMOHis_ncoI_f	catgccatgggccatcatcatcaccatcattcggactcagaagtcaatcaag

3.1.11 Plasmids

Table 15: Plasmids used in this work.

Plasmid name	Description	Source
pVP009	modified pCOLA Duet, Kan^R, N-His6	Own lab
pVP008	modified pCOLA Duet, Kan^R, N-Strep	Own lab
pPS018	modified pCOLADuet, Kan^R, N-His6_TEV_3xflag	Own lab
pPS041	modified pCOLADuet, Kan^R, N-His6_SUMO_TEV	Own lab
pPS001	pET28derived, InvD_P1737-N1976, Kan^R, N-His6-Thrombin	Prof. Petra Dersch's laboratory, HZI
pPS003	pCOLADuet (pVP009), InvD_G1640-N1976, Kan^R, N-His6-TEV	This study
pPS004	pCOLADuet (pVP009), InvD_P1838-N1976, Kan^R, N-His6-TEV	This study
pPS005	pCOLADuet (pVP008), InvD_G1640-N1976, Kan^R, N-Strep-TEV	This study
pPS016	pCOLADuet (pVP009), InvE_A2547-L2795, Kan^R, N-His6-TEV	This study
pPS017	pCOLADuet (pVP009), InvE_G2448-L2795, Kan^R, N-His6-TEV	This study
pPS019	pCOLA Duet (pPS018), InvD_G1640-N1976, Kan^R, N-His6_TEV_3xflag	This study
pPS020	pCOLA Duet (pPS018), InvD_G1640-G1839, Kan^R, N-His6_TEV_3xflag	This study
pPS021	pCOLA Duet (pPS018), InvA_P500-I985, Kan^R, N-His6_TEV_3xflag	This study
pPS022	pCOLA Duet (pPS018), InvE_G2448-L2795, Kan^R,	This study

	N-His6_TEV_3xflag	
pPS027	pCOLA Duet (pVP009), YadA_E26-E253, Kan^R, N-His6-TEV	This study
pPS029	pCOLADuet (pVP008), InvD_G1640-N1839, Kan^R, N-Strep-TEV	This study
pPS031	pCOLADuet (pVP009), InvE_G2448-L2795, Kan^R, N-Strep-TEV	This study
pPS032	pCOLADuet (pVP009), InvE_G2448-L2795, Kan^R, N-His6-TEV (Surface entropy mutation at aa 2695,2696)	This study
pPS033	pCOLADuet (pVP009), InvE_G2448-L2795, Kan^R, N-His6-TEV (Surface entropy mutation at aa 2646)	This study
pPS034	pCOLADuet (pVP009), InvE_G2448-L2795, Kan^R, N-His6-TEV (Surface entropy mutation at aa 2505, 2506)	This study
pPS035	pCOLADuet (pVP009), InvE_G2448-L2795, Kan^R, N-His6-TEV (Surface entropy mutation at aa 2648)	This study
pVK14 /pPS036	pET28 derived, RovC_R2-L247, Kan^R, N-His6-SUMO	Prof. Petra Dersch's laboratory, HZI
pPS037	pCOLA Duet (pVP009), RovC_D40-L247, Kan^R, N-His6-TEV	This study
pPS038	pCOLA Duet (pVP009), RovC_S80-L247, Kan^R, N-His6-TEV	This study
pPS042	pCOLA Duet (pPS041), RovC_M1-L247, Kan^R, N-His6-SUMO-TEV	This study
pPS047	pCOLA Duet (pPS041), RovC_M1-L247, Kan^R, N-His6-SUMO-TEV (Surface entropy mutation at aa 181,182,183 aa)	This study
pPS048	pCOLA Duet (pPS041), RovC_M1-L247, Kan^R, N-His6-SUMO-TEV (Surface entropy mutation at aa 233, 234 aa)	This study
pPS085	pCOLA Duet (pPS041), RovC_M1-L247 (S219E/A220E), Kan^R, N-His6-SUMO-TEV	This study
pPS092	pCOLA Duet (pPS041), RovC_M1-L247 (I150P/Y151P), Kan^R, N-His6-SUMO-TEV	This study

pPS093	pCOLA Duet (pPS041), RovC_M1-L247 (I150P), Kan^R, N-His6-SUMO-TEV	This study
pPS099	pCOLA Duet (pPS041), RovC_M1-L247 (R229E), Kan^R, N-His6-SUMO-TEV	This study
pPS100	pCOLA Duet (pPS041), RovC_M1-L247 (K56E), Kan^R, N-His6-SUMO-TEV	This study
pPS102	pCOLA Duet (pPS041), RovC_M1-L247 (R225E), Kan^R, N-His6-SUMO-TEV	This study
pPS107	pCOLA Duet (pPS041), RovC_M1-L247 (K211E/K215E), Kan^R, N-His6-SUMO-TEV	This study

3.1.12 Laboratory equipments

The technical equipments used in this study are given in Table 16.

Table 16: Technical equipments.

Equipment	Manufacturer
ÄKTApurifier	GE Healthcare
BP2100S, I2000D(Balance)	Sartorius AG
BandelinSonopuls HD 200(Ultrasonic Homogenizer)	BANDELIN electronic GmbH & Co. KG
CO8000 Cell Density Meter	Biochrom Ltd.
EmulsiFlex-C3(Homogenizer)	Avestin Inc.
Fujifilm scanner	Fujifilm Co-Kamera LAS-3000
Formulator(Liquid handling system)	Formulatrix Inc.
Gel Logic 212 Pro Gel documentation system	Carestream Health Inc.
Heraeus microbiological incubator	Thermo Fisher Scientific Inc
HerasafeMultifuge X3R(Centrifuge)	Thermo Fisher Scientific Inc
Heraeus Fresco 21(Centrifuge)	Thermo Fisher Scientific Inc
Honeybee 961 (Protein crystallization robot)	Digilab Inc.
LSM510Meta Confocal microscope	Carl Zeiss
Monolith NT.115	Nanotemper Technologies GmbH
Multitron Standard (Incubation shaker)	Infors AG
Mini-PROTEAN(SDS-gel running system)	Bio-Rad Laboratories

Milli-Q Advantage A10 (Water purification system)	Merck KGaA
Micro Star 17(R) (Centrifuge)	VWR International
Micromax-007 HF (X-ray generator)	Rigaku Corp.
NanoDrop 2000(UV-Vis spectrophotometer)	Thermo Fisher Scientific Inc
ND1000(UV-Vis spectrophotometer)	Thermo Fisher Scientific Inc
OryxNano (Protein crystallization robot)	Douglas Instruments Ltd
PowerPac 300/Basic (Power Supply)	Bio-Rad Laboratories
Peqgradient Thermocycler (PCR machine)	Peqlab
R-963 (Microwave)	Sharp K.K.
Rock Imager 182/1000	Formulatrix Inc
SteREO Discovery.V8 Stemi SV 8 (Stereomicroscope)	Carl Zeiss AG
SZ40 (Stereomicroscope)	Olympus
Sorvall RC6 Plus(Centrifuge)	Thermo Fisher Scientific Inc
Saturn 944+ R-AXIS IV++(X-ray detector)	Rigaku Corp.
SevenEasy pH meter	Mettler-Toledo International Inc.
Thermomixer	Eppendorf AG
Trans-Blot Turbo (Western blot transfer system)	Bio-Rad Laboratories
UV Transilluminator	Intas Science Imaging
UV cuvettes semimicro	BrandTech Scientific, Inc
Variomag Mono Direct (Magnetic stirrer)	H+P Labortechnik GmbH
XB6200D(Balance)	PrecisaGravimetrics AG
XS204(Balance)	Mettler-Toledo International Inc.

3.2 Molecular biology methods

3.2.1 Polymerase chain reaction

The gene of interest was amplified in a peqstar2x Thermocycler by using the PCR approach. Reaction components as listed in table 17 were assembled on ice and the PCR program listed in table 18 was run to selectively amplify the desired DNA.

Table 17: PCR reaction mix.

Component	Volume (µl)
Sterile water	31
$MgSO_4$ (25 mM)	3
dNTPs (2 mM)	5
Primer forward (10 pm/µl)	2
Primer reverse (10 pm/µl)	2
KOD hot Start DNA Polymerase buffer (10x)	5
Template DNA (10 ng/µl)	1
KOD Hot Start Polymerase	1

Table 18: PCR program steps.

Cycle step	Temperature (°C)	Time (s)	Number of cycles
Initial denaturation	95	120	1
Final denaturation	95	30	
Annealing	55	15	40
Extension	70	60	
Final extension	70	10	1
Hold	4	∞	1

To introduce mutations in the gene of interest, a two-step PCR was performed. Two independent PCR reactions were performed as described above to generate two primary PCR products in the first step. For the second step, 0.5 µl of both the purified PCR products from step 1 were used as a template and PCR reaction was performed using the flanking primers as described above.

3.2.2 DNA agarose gel electrophoresis and gel extraction purification

PCR products were visualized after loading them on 0.8% or 1.2% agarose gel (depending on the size of the product). 1x TAE was used as a running buffer. 9 volumes of sample were mixed with 1 volume of 10x DNA loading buffer. Agarose gels were stained with Roti gel stain (Carl Roth) according to the manufacturer´s instructions. Gels were visualized under UV light

using Kodak Gel logic 212 Imaging system. Gel extraction was performed with the QIAquick gel extraction kit according to the manufacturer's instructions.

3.2.3 Restriction digest

Cloning vector and PCR products were digested with the restriction enzymes from NEB in an appropriate NEB buffer according to the manufacturer´s instructions. 2-3 µg of PCR product (50 µl), 5 µl of Cut smart NEB buffer (10x), 1 µl of enzyme 1, 1 µl of enzyme 2 were mixed and incubated at 37°C for 1 h. To dephosphorylate the plasmid, digested vector was treated with 1 µl of CIP and the reaction was further incubated at 37°C for 1 h. Digested products were loaded on the agarose gel. Band of interest was excised and purified as described above. Concentration of purified DNA was determined by NanoDrop spectrometer.

3.2.4 Annealing of oligonucleotides

Small fragments generated on annealing of the oligonucleotides were used to insert tags/protease cleavage site in the vector or study DNA-protein interaction by electrophoretic mobility shift assay (EMSA) and microscale thermophoresis (MST).

To insert tags/protease cleavage site in the vector, complementary oligonucleotides were first phosphorylated at their 5' end followed by annealing. Reaction components as listed in table 19 were assembled. The mixture was incubated at 37°C for 30 mins followed by incubation at 65°C for 20 mins to heat inactivate the enzyme.

Table 19: Reaction mix to phosphorylate oligonucleotide.

Component	Volume (µl)
Sterile water	14
T4 DNA ligase buffer (10x)	5
Oligo (10 pm/µl)	30
T4 polynucleotide kinase	1

Phosphorylated oligonucleotides were diluted (8 µl oligonucleotide + 32 µl water) and 5 µl of each of the diluted complementary oligonucleotide were mixed. Program listed in table 20 was run to anneal the oligos.

Table 20: Program to anneal complementary oligonucleotide.

Cycle step	Temperature (°C)	Time (s)	Number of cycles
Step 1	95	120	1
Step 2	95 (-1°C/cycle)	60	85
Step 3	4	∞	1

For MST/EMSA, complementary oligonucleotides were mixed and annealed with the program listed in table 20.

3.2.5 Ligation

For ligation, 3 µl of digested and dephosphorylated plasmid, 1 µl of purified and digested insert (or 5 µl of phosphorylated and annealed oligos), 2 µl of T4 DNA ligase buffer (10x) and 1 µl of T4 DNA ligase were mixed and the reaction was kept at room temperature for 1 h. As a negative control, ligation without insert DNA was performed.

3.2.6 Transformation

For transformation, 100 µl of chemicompetent cells were thawed on ice, to which either 10 µl of the ligation reaction or 1 µl of the purified DNA was added and incubated on ice for 30 mins. It was followed by a heat shock at 42°C for 90 s with short cooling on ice (90 s). After addition of 1 ml of SOC media, reaction was incubated at 37°C for 45 mins at 600 rpm. The mixture was centrifuged for 1 min at 3000 rpm. 800 µl of supernatant was discarded and the cells were resuspended in the remaining 300 µl of media. 300 µl of the cells were plated on LB agar plate containing the appropriate antibiotics.

3.2.7 Colony PCR

Colony PCR was performed to determine the presence or absence of insert DNA directly from *E.coli* colonies. Either vector specific or combination of insert-vector specific primers were used.

Table 21: Colony PCR reaction mix.

Component	Volume (µl)
Sterile water	6
2x Red PCR Master Mix	10
Primer forward (10 pm/µl)	2
Primer reverse (10 pm/µl)	2

For the detection of positive clones, an isolated bacterial colony was touched with the sterile tip and dipped into the PCR reaction tube, with gentle twirling of the tip 3-4 times. PCR run listed in table 22 was performed. In parallel, 500 µl of LB media in Eppendorf was inoculated as pre-culture with the same tip and tubes were incubated at 37°C for 1.5 h.

Table 22: Colony PCR program steps.

Cycle step	Temperature (°C)	Time (s)	Number of cycles
Initial denaturation	95	300	1
Final denaturation	95	30	
Annealing	55	15	25
Extension	70	70	
Final extension	70	10	1
Hold	4	∞	1

The presence of insert was determined by loading the PCR product on 0.8% agarose gel along side a DNA size marker. If the clones were positive, 8-50 ml of culture was inoculated with the pre-culture from Eppendorf tube and incubated overnight in a shaker at 37°C.

Plasmid DNA was isolated for sequencing using QIAprep Spin Miniprep kit according to the manufacturer´s instructions.

3.2.8 Sequencing of DNA

600 ng of isolated plasmid mixed with 2 µl of sequencing primer (forward or reverse) was sent to MWG-Biotech or in-house sequencing service to obtain the sequence of interest. DNA sequences were analyzed by Chromas and Serial Cloner softwares.

3.2.9 Storage of bacteria

For permanent storage of bacterial clone, 500 µl of freshly grown overnight culture was mixed with 500 µl of 86% glycerol. Glycerol stocks were stored at -80°C.

3.3 Protein expression and purification

3.3.1 Small scale test-expression of recombinant genes in *E. coli*

Single bacterial colony of *E.coli* Rosetta2 (DE3) harboring protein of interest was used to inoculate 10 ml LB culture supplemented with the appropriate antibiotics. Culture was grown overnight in a shaker at 37°C. 300 ml of LB media was inoculated with the overnight culture to optical density at 600 nm (OD_{600}) of about 0.05 with the suitable antibiotics and the culture

was grown at 37°C until the OD_{600} reached 0.6-0.7. The culture was split into six small cultures with the volume of 40 ml each. To determine the optimum condition of expression of the protein of interest, a variety of parameters such as concentration of inducer and post induction incubation temperature were optimized. Cultures were induced with two different Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Carbosynth) concentrations (50 µM and 200 µM). The expression was tested at 16°C/18°C/20°C, 25°C and 37°C and the bacterial pellet was collected after 3 h and overnight incubation by taking 10 ml of cells (per condition) and centrifugation at 4000 rpm for 10 mins at 4°C. Bacterial cells were resuspended in 1 ml of lysis buffer. Cells were lysed by using ultrasound sonication (3x for 10sec @ 50% MS73D) followed by centrifugation at 13000 rpm at 4°C for 30 mins. Pellet was boiled in 30 µl of SDS-loading buffer to analyze the samples on SDS gel. Also 20 µl of supernatant was taken from each sample and mixed with 20 µl of SDS loading buffer. Cleared lysate was incubated with 20 µl of beads (depending on the affinity tag fused to the protein) and allowed to incubate on a rotating platform for 30 mins at 4°C. Beads were washed twice with 1 ml of washing buffer and SDS loading buffer was added to the beads. Beads sample along with the supernatant and pellet samples were analyzed on 15% SDS gel. The expression condition that yields the most soluble protein was used for large scale expression and purification.

3.3.2 Expression of recombinant genes in preparative scale

4 L of TB or dYT medium (supplied with suitable antibiotics) was inoculated to an OD_{600} of 0.05 with the overnight culture and it was grown at 130 rpm, 37°C until the OD_{600} reached 0.7. Expression was induced by adding IPTG and changing the incubation temperature according to the identified conditions in the test-expression. After incubation, cells were harvested by centrifugation at 5000 rpm for 15 mins at 4°C.

For expression of Se-Met derivative protein, transformed cells were first grown in 2x 500 ml of LB medium (supplied with suitable antibiotics) at 37°C, 130 rpm overnight. 500 ml of fresh LB medium (supplied with suitable antibiotics) was added to each of the overnight culture and cells were grown for another hour at 130 rpm, 37°C. Cells were harvested by centrifugation at 6000 rpm, 4°C for 5 mins and washed twice with Minimal Medium (by centrifugation at 6000 rpm, 5 min at 4°C). Pellet was resuspended in Minimal Medium. With this suspension, 4 L of Se-Met Medium was inoculated to the OD_{600} of 1.0, 50 mg/L of selenomethionine (Acros Organics) was added and protein production was induced by adding IPTG. Induction was performed under the conditions that were obtained in the small-scale test-expression. After overnight incubation, cells were harvested by centrifugation at 5000 rpm for 15 mins at 4°C.

3.3.3 Preparative lysis of bacterial cells

Bacterial cells were resuspended in the lysis buffer and lysed by homogenizing using Avestin homogenizer (2 cycles, air pressure 50-60 psi, homogenizing pressure 16000 psi). Lysate was cleared by centrifugation at 16000 rpm at 4°C for 60 mins.

3.3.4 Affinity chromatography

Affinity chromatography is based on the specific and reversible interaction between the molecules. A specific ligand is immobilized to a chromatography matrix. Complex mixture is then passed over the column. Molecules that have high specificity towards the ligand will bind while the non-binding components are removed using a wash buffer that will not interfere with the binding interaction between the ligand and the protein of interest.

Ni-affinity chromatography

Polyhistidine-tags are often used for affinity purification of recombinant proteins expressed in *E. coli* and other prokaryotic expression systems. The sample is incubated with affinity media such as Ni Sepharose. These matrices coordinate Ni^{2+} through four of the six coordination sites while the two remaining sites are exposed to interact with the polyhistidine tag. The protein mixture is passed through the matrix containing bound Ni^{2+} ions. Imidazole side-chain present in histidine will interact with the Ni^{2+}. Non-specific binding can be reduced by using low concentrations (1-10 mM) of imidazole in the binding buffer. The resin is then washed with the buffer containing slightly higher concentrations of imidazole to remove proteins that do not specifically interact with the matrix-bound Ni^{2+} ions; subsequently histidine-tagged proteins are specifically eluted with imidazole which competes with the polyhistidine-tag for binding to the resin.

To purify Histidine-tagged proteins, cleared lysate (after homogenization and centrifugation) was loaded onto 3 ml Ni-sepharose gravity column, which was pre-equilibrated with 50 ml of Ni-NTA lysis buffer. Ni-NTA resin (GE Healthcare) was allowed to incubate with the protein for 1 h at 4°C with gentle shaking. The Ni-NTA column was washed with 60 ml of Ni-NTA wash buffer and protein was eluted in Ni-NTA elution buffer.

After elution of the protein of interest, beads were washed using 2 CV 6M Guanidinium Hydrochloride, followed by washing with water and storage in 20% ethanol.

Strep affinity chromatography

It is based on the specific and reversible interaction between the strep tag-II (WSHPQFEK) and Strep-Tactin, an engineered version of naturally occurring tetrameric bacterial protein Streptavidin that has high affinity towards biotin. The short 8 aa tag can be placed at the N terminus or C terminus. The tagged protein binds to the immobilized Strep-Tactin during

affinity purification. After a short washing step, protein is eluted by addition of D-Desthiobiotin that competes for the biotin binding pocket (Schmidt and Skerra, 2007).

To purify the strep-tagged proteins, cleared lysate (suspended in strep lysis buffer) was loaded onto equilibrated 4 ml Strep-Tactin Sepharose beads gravity column. On incubation of lysate with the beads, column was washed with strep wash buffer and protein was eluted in elution buffer containing 3 mM D-Desthiobiotin (iba-2-1000-005).

After elution of the protein of interest, beads were washed using 3 CV 0.5 M NaOH and stored in 50 mM TRIS pH 8.0, 200 mM NaCl.

3.3.5 Removal of affinity tag

If there is a need to remove the fused tag, elution fractions were pooled and concentrated to volume of 10 ml using viva-spin concentrators (GE Healthcare) with 10000 molecular weight cut-off (MWCO). Concentrated sample was dialyzed in a 3500 Da molecular weight dialysis tube (Thermoscientific) in TEV cleavage buffer and the tag was removed by overnight TEV cleavage (ratio 1 mg to 30 mg protein) at 4°C for 15 h. The fusion tag, TEV protease and uncleaved protein were removed by passing the dialyzed sample again through Ni/strep column.

3.3.6 Concentration of proteins by ultrafiltration

Protein solutions were concentrated using viva-spin concentrators with 10000 MWCO. Centrifugation is carried out at 4000 rpm, 4°C with gentle pipetting of the protein solution during concentration to prevent protein precipitation by accumulation of highly concentrated protein at the bottom of the filter over the membrane. Proteins larger than the filter cut-off size will be retained by the filter.

3.3.7 Size exclusion chromatography

Size exclusion chromatography (SEC) is a method of separating protein mixtures on the basis of their size and shape. SEC matrix consists of pores of different sizes. The separation of the molecule depends on the ability of the protein to enter the channels in the porous beads. Small molecules easily enter the porous beads while large molecules are restricted and pass between the beads. So, all molecules with size larger than the pore size are completely excluded to enter the porous channels and elute together.

Protein was loaded onto the column that was pre-equilibrated with 1CV of filtered (0.2 µm whatman membrane filter-GE Healthcare) and degassed gel filtration buffer. Fractions containing the protein were analyzed on SDS-PAGE. Fractions that contain the protein of interest were pooled, concentrated, aliquoted and flash frozen in liquid nitrogen and kept at - 80°C for further use.

3.4 Protein characterization and modification

3.4.1 Protein concentration according to nanodrop

Protein concentration was determined by nanodrop at 280 nm using the default settings – E 0.1 % (1 mg/ml). Spectrophotometric determination of protein concentration depends on Beer Lambert´s law which gives the relation between the absorption of light and the amount of absorbing substance present. It is expressed as-

$A = \varepsilon_{280} \times c \times l$ Equation 1

where A is the absorbance, c is the concentration of sample in mol L^{-1}, l is the length of path light in cm and Ɛ is the molar extinction coefficient or molar absorptivity (expressed in L mol^{-1} cm^{-1})

Protein concentration obtained by nanodrop was divided by the Absorbance (0.1%= 1 g/L) (which is calculated as mass extinction coefficient/Molecular weight) obtained from EXPASY Protparam (Gasteiger et al., 2005).

Table 23: Molecular size and Abs (0.1%=1g/L) of the proteins used in this study.

Protein name	Molecular weight (kDa)	Abs (0.1%=1g/L)
InvD G1640-N1976	34.5	0.634
InvD P1737-N1976	25.0	0.831
ΔAD-InvD	19.7	0.151
InvA P500	51.2	1.022
InvE	36.4	0.944
ΔAD-InvE	23.2	0.429
YadA	23.0	0.194
RovC M1-L247	28.7	1.960

3.4.2 Denaturing SDS gel electrophoresis

It is a technique to separate components of a protein mixture based on their size. It is commonly used to determine purity of the protein, protein subunits and protein quantitation. Protein is incubated with SDS and DTT (SDS loading buffer) at 95°C for 5 mins. Under these conditions protein becomes fully denatured and dissociated from each other. In addition, SDS binds to the protein on a weight basis (1.4 g SDS/g of protein or about 1 SDS molecule per two amino acids). The SDS-protein complex forms a rod with its length proportional to the molecular weight of the protein. The complexes are then separated on the bases of their size within resolving gel matrix, where small protein-SDS complex moves faster while larger complexes move slowly though the matrix.

Electrophoresis was performed in 1x SDS running buffer at a constant voltage of 200 V for 40 mins. Protein standard is applied to compare the molecular weight. After completion of electrophoresis, gels were stained with the staining solution. Gels were destained by incubating in the destaining solution for 15-20 mins or boiling the gel in water for 5-7 mins.

3.4.3 Western blot

Western blot is a commonly used technique to detect the presence of a specific protein or a tag in a complex mixture with high sensitivity. For western blotting, SDS gel was first washed with water, and then incubated in transfer buffer for 10 mins. Polyvinylidene difluoride membrane (PVDF) (EMD Millipore-IPVH00010) was activated by 100% methanol for 5 s, washed with water and then incubated in transfer buffer for 10 mins. Whatman paper was also incubated in transfer buffer. Proteins were transferred by semi dry transfer method for 30 min at 25 V, 1 A. Next, the membrane was washed with TBS-T and then blocked with 5% skim milk (in TBS-T) for 1 h at room temperature. Membrane was incubated with the primary antibody (1:2000 diluted in TBS-T) for 2 h. Again, the membrane was washed with TBS-T. The membrane was incubated with the secondary antibody for 1 h, and washed with TBS-T before detection. In case of HRP conjugated antibody, membrane was developed using Lumi-light Western Blotting Substrate (Roche-12015200001) and analyzed by Fujifilm LAS-3000 Imager. In case of AP conjugated antibodies, membrane was developed with Alkaline Phosphatase staining solution.

3.4.4 Limited proteolysis

Limited proteolysis was performed to either search for the stable subdomains of the protein of interest or test protein's stability in presence of different proteases. This method is based on the susceptibility of conformationally flexible and exposed regions of protein to proteases whereas the protein core remains resistant to proteases (Dong, 2007). Limited proteolysis was performed using Hampton Proti-Ace kit I and II according to manufacturer´s instructions. In brief, 0.01 mg/ml of protease solution was mixed with 1 mg/ml or 5mg/ml of protein in equal volume and incubated at room temperature, 37°C or 4°C for different time points. Reaction was stopped by addition of SDS-PAGE sample loading dye. Subsequently samples were analyzed by SDS-PAGE and/or MS.

3.4.5 Mass spectrometry

For the identification of protein bands, band of interest was excised from blue silver stained or Coomassie stained SDS gel and subjected to mass spectrometry (matrix-assisted laser desorption ionization) analysis performed by Mass Spectrometry platform at HZI using Ultraflex TOF/TOF (Bruker Daltonics). In-gel tryptic digestion was performed and the size of

the resulting fragments was determined by mass spectrometry. Peptide sequences thus obtained were searched in protein database to identify the protein.

3.4.6 Protein labeling

The accessible primary amine residues of the protein were covalently labeled with Cyanine-5 NHS ester (GE Healthcare) in PBS according to the manufacturer's instructions. 20 µM (100 µl) of protein in 1xPBS was mixed with 60 µM (100 µl) dye in DMSO. This mixture was incubated at room temperature in dark for 30 mins. Excess dye was removed by buffer exchange column (illustra NAP-5 columns GE Healthcare). The protein:dye ratio was determined from the protein absorption at 280 nm and the Cy5-absorption at 647 nm using an extinction coefficient of Cy5 of 250000 L mol^{-1} cm^{-1}.

3.4.7 Size Exclusion Chromatography-Multi Angle Light Scattering

In Multi Angle Light Scattering (MALS), a beam of light is focused on the sample in solution. Light scattered by the sample is detected at various angles at the same time. The intensity of the light scattered is directly proportional to the molecular mass of the macromolecule. Size exclusion chromatography (SEC) combined with MALS is used to determine the exact molecular weight of the macromolecule in solution regardless of its shape and conformation which cannot be determined using UV detector and SEC alone.

For the molecular mass determination of RovC, an inline SEC (Akta PURE, GE Healthcare) coupled with a MALS detector (miniDAWN TREOS, Wyatt Technology) and a differential refractometer (RI) (Optilab T-rEX, Wyatt Technology) was used (Prof. Teresa Carlomagno's laboratory). 2.8-3.3 mg/ml of RovC was loaded on Superdex 200 10/300 GL column prequilibrated with 100 mM TRIS pH 8, 500 mM NaCl and 5 mM DTT. Molecular weight of the eluted protein was analyzed by ASTRA 6 software (Wyatt Technology) by recording RI and MALS signals.

3.4.8 Lysine methylation

Lysine methylation is an inexpensive, fast and simple process to enhance crystallization by methylation of the surface exposed Ɛ-NH_2 group of lysines, which changes the protein's biophysical properties. The protocol is adapted from (Walter et al., 2006). RovC (5 mg/ml_1.5 ml) was dialyzed against lysine methylation dialysis buffer (2 L) at 4°C, overnight. RovC was diluted to a concentration of 1 mg/ml and 20 µl 1 M dimethylamine-borane (Sigma Aldrich-180238) and 40 µl 1M formaldehyde (37% formaldehyde solution, Carl Roth) per ml of protein solution was added. Sample was gently mixed and incubated at 4°C for 2 h. Again, 20 µl 1M dimethylamine-borane and 40 µl 1M formaldehyde per ml protein solution was added and the sample was gently mixed and incubated for 2 h at 4°C. Finally, 10 µl 1 M

dimethylamine-borane per ml of protein solution was added and the sample was incubated at 4°C, overnight. Protein was centrifuged at 4°C, 13000 rpm for 10 mins to remove precipitated protein. Protein was purified on gel filtration column (S200 10/300), pre-equilibrated with lysine methylation gel filtration buffer.

3.4.9 Surface entropy reduction

Protein crystallization is a thermodynamic process and depends on the entropy of the system.

According to Gibb´s Helmholtz equation,

$\Delta G = \Delta H - T(\Delta Sp + \Delta Ss)$ Equation 2

Where *G*, *H*, *T*, Sp and Ss are Gibb's free energy, enthalpy, temperature, protein entropy and solvent entropy, respectively.

Equation 2 represents the free energy change that drives crystallization. A negative ΔG is a prerequisite for successful crystallization. When protein crystallizes, there is loss in entropy for the protein molecules making ΔSp negative, but the release of water molecules during protein crystallization contributes to entropy gain of the solvent making $(\Delta Sp + \Delta Ss)$ highly positive, favoring crystallization by an overall negative contribution of $-T(\Delta Sp + \Delta Ss)$ to ΔG.

In the Surface Entropy Reduction approach, highly entropic residues on the surface of protein (mostly Lys, Gln, Glu) are replaced by less entropic residues (mainly Ala). Due to these mutations, the total entropy loss of the mutant protein is smaller compared to its native form. With constant ΔSs , the total entropy difference in equation 2 becomes larger making $-T(\Delta Sp + \Delta Ss)$ and therefore also ΔG more negative, which could enhance the crystallization of the protein (Derewenda and Vekilov, 2006). Highly entropic residues were predicted by Surface Entropy Reduction prediction software-UCLA MBI (SERp) (Goldschmidt et al., 2007).

3.5 Protein interaction

3.5.1 Pull downs

Pull down assays were performed to analyze protein–protein interactions. 5-20 µg of tagged bait protein in 1x TBS (or 1x PBS) was first immobilized to 20 µl affinity matrix (specific for the tag of the bait). On incubation, beads were washed twice (centrifugation at 2000 rpm, 4°C for 2 mins) with 1x TBS (or 1x PBS) to remove any unbound protein. Beads were then incubated with 1 ml of homogenized extract of mouse tissue or with 6 µg of purified protein in 100 µl of 1xTBS (or 1x PBS). Following incubation at 4°C (o/n with homogenized tissue extract or 2 h with purified protein), with gentle shaking, beads were washed twice (or 3x) with 1xTBS (or 1xPBS) and protein was eluted. Wash, elution fractions and beads were

analyzed by SDS-PAGE analysis and stained by blue silver stain (Candiano et al., 2004)/Coomassie blue staining solution.

3.5.2 Biopanning

Phage display panning was performed by Saskia Helmsing in Prof. Michael Hust's laboratory.

(Bio-) panning is the procedure of selecting binding partners from phage display library. In this affinity-driven selection procedure, protein of interest is incubated with phages, washing is done to remove unbound phages and bound phages are eluted. This process is repeated for several cycles to enrich the target protein with specifically bound phages (Hammers and Stanley, 2014)(Adams et al., 2014).

HAL9 and HAL10 scFv libraries with size of $1.04*10^{10}$ and $4.45*10^{9}$ respectively were used for panning (Kügler et al., 2015)**.** The panning and production of soluble scFv was performed according to (Frenzel, Kügler, Wilke, Schirrmann, & Hust, 2014). In brief, 5 µg of strep-tagged InvD1640 was coated on microtiter plate (U96 PP 0.5 ml, Greiner, Frickenhausen, Germany) and incubated with HAL10 library/HAL9+HAL10 library. After a few washings, bound phages were eluted with trypsin. The eluted phages were used to infect *E.coli* culture which was then co-infected with the helper phage and grown overnight. Phage particles were precipitated in PEG/NaCl and then used for the next round of panning. This process was repeated for three rounds and plasmids of isolated clones were purified and inserts sequenced.

Production of soluble scFv

For the production of soluble scFv, clones were picked from the third round of panning, inoculated in dYT–GA media and incubated overnight at 30°C at 250 rpm. Phosphate–buffered dYT-GA media was inoculated with the overnight culture and grown for 2 h at 37°C and 800 rpm. Cells were harvested and pellet was suspended in 2xYT supplemented with 100 µg/ml ampicillin, 100 mM sucrose and 50 µM IPTG and incubated at 30°C and 800 prm overnight. Bacteria were pelleted and supernatant was used for ELISA.

Enzyme linked immunosorbent assay

A 96 well microtiter plate was coated with 200 ng of InvD 1640 or ΔAD-InvD (in PBS pH 7.4). As a control, four wells were also coated with lysozyme. After overnight incubation at 4°C, plates were blocked with MPBST for 1 h followed by three washings with PBST. 100 µl of scFv containing supernatant were added to the wells and incubated for 1.5 h at room temperature. Wells were washed three times with PBST and incubated with 9E10 mouse anti-myc antibody (1:50 dilution in MPBST) (Yumab, Braunschweig) for 1.5 h at room temperature. Plates were washed again as described earlier. For detection, goat anti-mouse

IgG-HRP coupled antibody (1:1000 dilution in 2% MPBST) (Sigma Aldrich-A0168) was added to the wells and incubated for 45 mins at room temperature followed by addition of TMB (3,3′,5,5′-Tetramethylbenzidine) and absorbance was measured at 450 nm (620 nm reference) by ELISA plate reader (Tecan, Crailsheim, Germany).

3.5.3 Beads assay

Coating of latex beads with the protein

30 µl suspension of amine modified polystyrene, fluorescent yellow-green latex beads (L1030; Sigma-Aldrich) were added to 1 ml of 1x PBS. After one washing step with 1x PBS, beads were suspended in 1 ml coupling buffer in the absence of light. After centrifugation at 12000 rpm for 5 mins, beads were resuspended in 100 µl of coupling buffer and 100 µl of 0.6-1 mg/ml InvA, BSA or InvD1640 were added to it and incubated at 37°C for 1 h. Then 500 µl of coupling buffer was added and sonicated for 10 seconds. To the coated beads 500 µl of 30 mg/ml BSA in coupling buffer was added followed by 1 h incubation at 37°C followed by washing with 1x PBS in 0.2 mg/ml BSA. Coated beads were resuspended in 1 ml 1xPBS 0.2 mg/ml BSA and stored at 4°C until use.

Cultivation of eukaryotic cells

Human Epithelial cells (HEp-2 cells) were cultivated in 75 cm^2 cell culture flasks (Greiner) with 20 ml of HEp2 medium at 37°C, 89% humidity and 5% CO_2 in a cell incubator (Heracell 150, Prof. Petra Dersch's laboratory). HEp-2 cells were passaged every second day. Cells were washed with 1x PBS and treated with 2 ml trypsin-EDTA solution for 5 mins at 37°C. Trypsinization was stopped by the addition of 8 ml culture medium. 1 ml of this suspension was diluted with 19 ml of medium and transferred to a new flask.

Cell density measurement

Cell density was measured by Neubauer counting chamber. 10 µl of resuspended cell suspension was mixed with 90 µl of trypan blue and loaded onto the counting chamber. Trypan blue stains dead cells blue while live cells do not take up the dye. Counting chamber was observed under the microscope and live cells were counted in each of the four squares. The cell number per ml was calculated by the equation-

$$Concentration\ (cell\ number\ per\ ml) = \frac{number\ of\ cells * 10{,}000}{number\ of\ cells * dilution\ factor} \qquad \text{Equation 3}$$

The cells were resuspended in the appropriate volume of the culture medium to reach the desired cell density and seeded into a new flask.

Cell adhesion assay

For cell adhesion assay, $5*10^4$ HEp-2 cells were seeded and grown overnight in individual wells of 12-well cell culture plates. Cells were washed three times with PBS and incubated in HEp-2 binding buffer. Approximately $2*10^6$ beads were added to the cells and cell culture plate was centrifuged for 5 mins at 1000 rpm. Cells were incubated for 1 h at 25°C to check for adhesion. After 1 h incubation, cells were washed 3 times with PBS and fixed with 2% paraformaldehyde (Sigma-Aldrich-P6184). After 10 mins post-incubation at room temperature, cells were washed once with PBS and fixed cells were stained with 4´, 6-diamidino-2-phenyl indole (DAPI; Roth-6335.1) to visualize the nuclei of the cells. After another 4 mins incubation at room temperature, cells were again washed with PBS and after addition of 200 µl of PBS, cells were visualized with a Zeiss Axiovert II fluorescence microscope and the AxioVision Software (Zeiss).

3.5.4 Flow cytometry

Flow cytometry experiments were performed by Jörn Pezoldt in Prof. Jochen Hühn's laboratory.

Single cell suspensions were obtained by mechanical disruption of the organs through a 100 µm cell strainer (BD, Franklin Lakes, NJ, USA). To eliminate erythrocytes, spleen cells were incubated for 3 mins in erythrolysis buffer (7.8 mM NH_4Cl, 10 mM $KHCO_3$, 100 mM EDTA). Dead cells were identified using Live/dead fixable blue dead cell stain kit (Invitrogen, Waltham, MA, USA). To exclude unspecific binding of antibodies, samples were incubated for 15 mins with anti-CD16/CD32 (BioXCell, West Lebanon, NH, USA) and rat IgG (Jackson ImmunoResearch, Baltimore, PA, USA). 3xflag tagged InvD1640 and AD-InvD were used at 100 µg/ml and staining was performed together with the antibodies staining (listed below table 5) for 15 mins at 4°C in PBS containing 0.2 % (w/v) bovine serum albumin (BSA, Sigma-Aldrich) and 2 mM Ca^{2+} (Roth, Karlsruhe, Germany). Flow cytometry was performed using LSR Fortessa flow cytometer with Diva software (BD), and data were analyzed with FlowJo software (TreeStar, Ashland, OR).

3.5.5 Lipid binding assay

To assess lipid binding of InvD1640, a protein-lipid assay was performed using Membrane Lipid Strips (P-6002; Echelon Biosciences) according to the manufacturer's recommendations. Briefly, membrane was blocked in 3% (w/v) fatty acid-free BSA in TBST for 1 h at room temperature in the dark followed by 1 h incubation with 3xflag-tagged InvD1640 (1.0 µg/ml in TBST) at room temperature with gentle agitation. After washing the membranes three times over a 1 min period in TBST, they were incubated for 1 h with 0.8 µg/ml rabbit anti-flag polyclonal antibody. Membranes were washed as before and

incubated for 1 h with anti-rabbit-HRP conjugate followed by washing as done before and detection by Echelon´s K-TMBP.

3.5.6 Glycan binding assay

Binding of 3xflag tagged InvD was analyzed by the Consortium for Functional Genomics (CFG) Protein-glycan interaction core. Binding of the protein to 609 different glycans on a glass slide (version 5.2) was analyzed at a concentration of 200 µg/ml. Protein binding was detected by anti-flag antibody and fluorescently labeled anti-rabbit IgG 488 antibody (supplied by the Consortium). The array consists of six replicates of each glycan. The highest and lowest point from each set of six replicates has been removed so the average is of four values rather than six to eliminate some of the false hits that contain a single very high or low point. 20 mM TRIS-HCL pH 7.4, 150 mM sodium chloride, 0.05% Tween 20 and 1% BSA was used as the binding buffer. As a washing buffer, 20 mM Tris-HCL pH 7.4, 150 mM sodium chloride, 2 mM calcium chloride, 2 mM magnesium chloride, 0.05% Tween 20 and 1% BSA was used.

3.5.7 Microscale thermophoresis

Microscale thermophoresis is based on the movement of molecules in a temperature gradient. This process is dependent on the size, hydration shell or charge of the molecule which has an influence on the Soret coeffecient (S_{Ti}).

$$S_{Ti} = \frac{D_{Ti}}{D_i} \qquad \text{Equation 4}$$

Where D_{Ti} is the thermophoretic mobility or thermal diffusion coefficient and D_i is the diffusion coefficient.

One of these parameters gets affected when a molecule binds to its interaction partner hence changing the Soret coeffecient. The Microscale thermophoresis technique detects these changes in the movement of the molecule upon binding to an interaction partner (Duhr and Braun, 2006; Jerabek-Willemsen et al., 2014).

In this method, one of the interaction partners is fluorescently labeled according to the protocol described in section 3.4.6. Constant amount of labeled protein was then titrated with the increasing concentration of the unlabeled partner, and the change in thermophoresis of fluorescently labeled molecule in presence of unlabeled partner is measured, which leads to the measurement of binding constant. 12 or 16 dilutions of unlabeled partner were prepared and mixed in a 1:1 ratio with the labeled protein. Samples were then filled in the standard treated capillaries. Infrared laser is focused on the capillary to create a temperature gradient of 2-6 K. The temperature gradient leads to depletion of molecules in that region which is quantified by Soret coefficient.

The experiments were performed on Monolith NT.115 at 20% LED with a MST power of 20–60 %; laser-on time was 30 s, laser-off time 5 s each. Curve fitting was done by MO.Affinity Analysis (x64) software (NanoTemperTechnologies) in Thermophoresis + T-jump mode.

3.6 X-ray crystallographic methods

3.6.1 Protein crystallization

To elucidate the structure of proteins by X-ray crystallography, protein crystals are required. For the formation of protein crystals, protein in solution is mixed with a precipitant solution to decrease the solubility of the protein which leads to supersaturation and hence crystal formation. Large supersaturation is required to overcome the activation energy required to form crystals. Protein crystallization can be explained by crystallization phase diagram (Figure 13). This is a two-dimensional solubility plot with X-axis showing precipitant concentration and protein concentration on Y-axis. The solubility curve is divided into two zones: the undersaturated and the supersaturated zone. In the undersaturation region protein is soluble and no crystal formation will take place. The supersaturation zone starts above the solubility curve and is further divided into three zones (Asherie, 2004):

1) metastable zone-spontaneous nucleation will not occur, crystal growth only occurs from pre-existing nuclei.

2) nucleation zone- supersaturation is large enough that crystal growth can take place.

3) precipitation zone- supersaturation is too large which leads to precipitation of the protein.

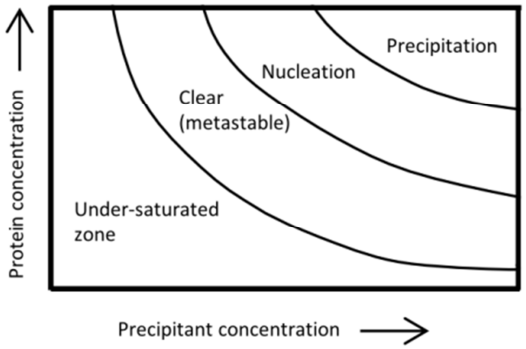

Figure 13: Crystallization phase diagram. Schematic representation of two-dimensional crystallization phase diagram.

As illustrated in the protein crystallization diagram, protein crystallization depends on protein as well as precipitant concentration, pH, ionic strength and additives. Apart from these factors, temperature has also been shown to be an important variable in protein crystallization. Protein solubility can decrease or increase with the temperature depending on the type of precipitant used, thus it influences nucleation and crystal growth. Crystallization also depends on the geometry of the set-up. Two experimental setups are commonly used in crystallization, namely hanging drop and sitting drop (Figure 14). In both methods, protein

solution is mixed with the precipitant in a 1:1 ratio and is allowed to equilibrate with the precipitant solution (reservoir solution). As the protein drop contains less concentration of precipitant as compared to the reservoir solution alone, there is a net transfer of water (by vapor diffusion) from the protein drop to the reservoir. Hence, the precipitant and protein concentration in the protein drop is steadily increased until equilibrium is established.

In case of hanging drop, a well is filled with 500 µl of reservoir solution. 500 nl of protein solution was mixed with 500 nl of precipitant solution on a cover slip which was flipped over the well. The contact area between the coverslip and the well was sealed with silicone grease. For this approach, 24 well plates (VDX Plate^{TM} HR3-108) were used.

In case of sitting drop, 96 intelli plates (Art Robbins Instruments, Hampton Research) were used. Screening was performed with Honey bee robot or drops were set manually. With honey bee, wells were filled with 60 µl of reservoir solution. The protein drop (200 nl) was mixed with reservoir solution (200 nl) and the plate was covered with transparent tape and placed in the RockImager for imaging of drops. For manual set-ups, 60 µl of reservoir solution was placed in the well and 400 nl of protein was mixed with 400 nl of reservoir solution.

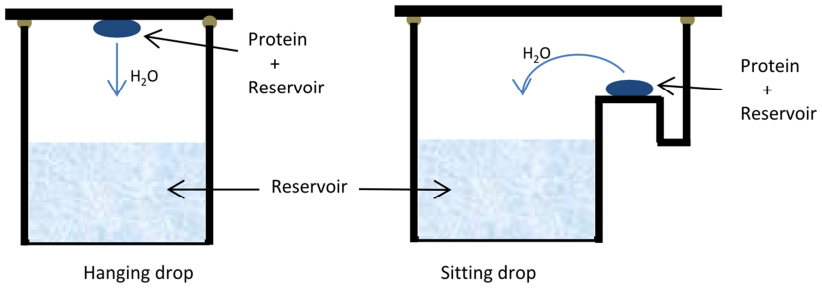

Figure 14: Hanging and sitting drop methods. Schematic representation of the hanging drop and sitting drop methods for protein crystallization.

For initial screening, different crystallization screens were used as mentioned in Table 10. Protein concentration ranging from 2 mg/ml-30 mg/ml was used in protein crystallization experiments and plates were incubated at 20°C and 4°C.

3.6.2 Cryoprotection

When the protein crystal reaches a certain size, it is harvested from the drop and flash frozen in liquid nitrogen, if the drop contains sufficient amount of cryoprotectant. If this is not the case, the crystal is first dipped in a drop of cryoprotectant and then quickly dipped into liquid nitrogen. Cryoprotectants prevent the formation of ice crystals as they act as an anti-freezing agent and prevent the mother liquor from freezing as amorphous ice. If there are ice crystals on the loop, it largely affects the quality of the diffraction. Cryoprotectants used in this study are PEG400, glycerol and sodium formate.

3.6.3 Structure determination by X-ray crystallography

When X-rays strike a crystal, they get diffracted upon interaction with the electrons of the atom. The diffracted rays consist of reflections of different intensities from sets of parallel planes (called Miller planes characterized by three Miller coordinates (h, k and l)) of atoms in a crystal.

According to Bragg´s law, for diffraction to occur, the following rule must be satisfied,

$n\lambda = 2dsin\theta$ Equation 5

where n is an integer, λ is the wavelength of the incident wave, d is the spacing between the atomic planes and θ is the incident angle (reflection angle) which is the angle between the incident beam and the plane. As the crystal consists of an infinite set of Miller planes, each plane will diffract X-rays differently. The interference can only be constructive when the path difference is equal to the integral number of wavelength (Figure 15). From the X-ray diffraction pattern that is generated by these planes, crystal parameters can be calculated.

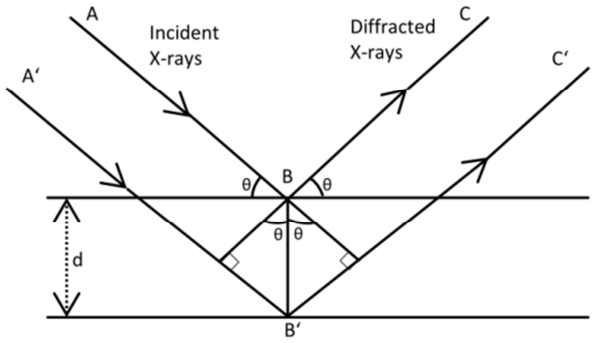

Figure 15: Schematic representation of Bragg's law. Two X-ray beams A and A' arrive in phase and strike the planes with an angle θ. The diffracted X-rays can only be constructive when the distance between the paths ABC and A'B'C' is an integral number of wavelength.

Intensities of the reflections can be calculated from the experimental data obtained from the X-ray experiment. The major step towards determination of a structure is to construct the electron density map, which is calculated by Fourier Transformation of the diffraction data. The function for electron density $\rho(x,y,z)$ is given as

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{-2\pi\cdot i(hx+ky+lz-\alpha_{hkl})} \quad \text{Equation 6}$$

Where V is the unit cell volume, $|F_{hkl}|$ is the structure factor amplitude for each reflection and α_{hkl} are the phases for each reflection.

To calculate the electron density, one needs the information of the intensity of the reflections, which is proportional to $|F_{hkl}|^2$, and the phases (α_{hkl}). Intensities are directly measured by the detector but the phase information is lost during the measurement. This is known as phase problem. In protein crystallography, phases can either be derived from the coordinates of structurally similar proteins (called Molecular Replacement (MR)) or experimentally by

determining the positions of heavy atoms that have been added to the protein (methods include single/multiple isomorphous replacement (SIR/MIR), single/multiple wavelength anomalous dispersion (SAD/MAD). Only SAD and MR will be discussed further in this work.

3.6.4 Molecular replacement

Molecular replacement (MR) is based on the assumption that the phases of a homologous protein (model) will provide the estimation of the real phases. In MR, the model is oriented and positioned in such a manner that it coincides with the crystallized target protein in the unit cell.

To determine the correct orientation and position of the model in the unknown unit cell, three rotation angles and three translational directions must be specified. For one molecule in the asymmetric unit, the MR problem becomes a 6D problem. To solve this problem, 6D problem is divided into two 3D problems. First rotation angles are calculated by the use of Patterson function, and then the oriented molecule can be placed in the unit cell by 3D translation function.

If Fourier transformation is applied only on the intensities, a Patterson map is obtained. This map gives the information about the interatomic distances within the molecule. This information can be used to solve the structure for small molecules, but as the number of the peaks in the Patterson function is directly proportional to the square of the number of atoms, it is not possible to solve the structure for large molecules. However, this information can be used to solve the orientation problem. Patterson function of the model is calculated for every possible orientation and is matched with the Patterson function of the unknown structure. Once the correct orientation is found, the same method is used to correctly determine the position of the molecule in the unit cell. Once the correct solution is obtained, reverse Fourier transformation is applied to calculate the phases which are further used to calculate the electron density map (Evans and McCoy, 2007).

3.6.5 Anomalous diffraction

If the incident X-ray energy is close to the transition energy of the used heavy metal atom, the phase and the intensity of the diffracted wave gets affected. Since some of the energy is used to bring the atom to an excited state, the intensity of the coherent scattering at this wavelength is reduced and also there is a phase change of the scattered X-rays. This effect is called anomalous scattering.

In the presence of anomalous scatterer, the scattering power of that atom at any wavelength becomes-

$$f = f_n + \Delta f' + i\Delta f''$$ Equation 7

Where f_n is the normal scattering factor (independent of wavelength), $\Delta f'$ is the real component that describes the amount by which the normal scattering is reduced and $\Delta f''$ is the imaginary component which is 90° out of the phase with the normally scattered radiation. These factors are wavelength dependent and vary with the wavelength used for the measurement. f'' has its maximum value at the peak wavelength, while at the edge wavelength f' is reduced to its lowest possible value.

Anomalous phasing is based on the violation of Friedel's law. According to Friedel's law, structure factor F_{hkl} and its Friedel mate F_{-h-k-l} have the same magnitude but opposite phases. In the presence of anomalous scatterers, F_{hkl} and F_{-h-k-l} have different magnitudes and also their phases are different giving rise to an anomalous difference which can be used to locate the anomalous scatterer. Once the position of the heavy atom is located in the unit cell by Patterson methods, phases can be calculated.

In SAD (single anomalous diffraction/dispersion), data is calculated at the peak wavelength of the anomalous scatterer. In multi wavelength anomalous dispersion (MAD), datasets at three different wavelengths are collected: first dataset is collected at the peak wavelength of the anomalous scatterer, second at the remote wavelength and third at the inflection wavelength (Taylor, 2003).

In this study, to solve the phase problem by SAD, heavy metals are introduced into the preformed crystals by soaking, for the determination of InvD structure. For InvE and RovC, Selenomethionine variants of the proteins were expressed, purified and crystallized. Data were processed with XDS software package (Kabsch, 2010). Phaser MR was used to perform Molecular Replacement (McCoy, 2006). Phaser EP was used for MR-SAD and SAD phasing (Read and McCoy, 2011).

3.6.6 Model building and refinement

After determination of phases, the calculated electron density is interpreted and model is built. To eliminate the conformational or stereo-chemical errors that remained after model building, it is necessary to carry out a crystallographic refinement. During refinement, several parameters such as B-factors, geometry and model coordinates are refined against experimental data to obtain the best fit between the observed and calculated structure factor amplitudes. The overall fit is calculated by the following equation

$$R = \frac{\sum_{h,k,l}\left||F_{obs}(h,k,l)|-|F_{calc}(h,k,l)|\right|}{\sum_{h,k,l}|F_{obs}(h,k,l)|} \qquad \text{Equation 8}$$

R or R_{work} is a useful factor to assess the progress of the model. However, it can be lowered by over-fitting or mis-fitting the data, but this would not improve the model. To avoid this problem, R_{free} was introduced (Brunger, 1992). R_{free} is calculated in the same way as R_{work}, but for R_{free} calculation 5% of the reflections are excluded from the structure refinement. If

there is an over-fit in the model, R_{free} will increase or will stop dropping, while R_{work} is further reduced. The initial model was built using Phenix Autobuild. Wincoot (Emsley et al., 2010) and Phenix_refine (Afonine et al., 2012) was used for iterative model building and refinement respectively.

3.7 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) reveals the structure of macromolecules in solution at a resolution of 10-20 Å. SAXS, like X-ray crystallography uses high energy X-rays to obtain structural information of the sample. In a SAXS experiment, an incident X-ray beam falls onto a thin glass capillary filled with protein solution. Scattered X-rays are recorded by the detector. As the molecules are randomly oriented in solution, averaging of these molecules leads to loss of information as compared to crystallography where the sample is in crystalline form. SAXS experiment gives information about the oligomerization state and also the organization of oligomers in solution. It is a useful technique to calculate the low resolution envelope of the molecule. SAXS is a contrast method; it depends on the difference in the electron density between the molecule and the solvent. Therefore, a buffer blank has to be measured and subtracted from the scattering curve of protein + buffer scattering curve to obtain the scattering curve of the pure protein.

The data obtained from SAXS experiments were processed using tools in the ATSAS software package. The net protein-scattering data was generated by subtracting scattering of buffer from the scattering of protein using PRIMUS (Konarev et al., 2003). The distance distribution function was calculated by GNOM (Svergun, 1992). 18-20 independent ab-initio models were built using DAMMIF (Franke and Svergun, 2009). These models were then averaged with DAMAVER package (Beaucage, 1995) to yield SAXS envelope. Fitting of the crystal structure with the experimental SAXS curve was performed using CRYSOL (Barberato et al., 1995). Using UCSF Chimera (Yang et al., 2012), the X-ray structure was fitted in the SAXS envelope.

SAXS data were measured for InvD1640 on the P12 beamline at PETRAIII EMBL Hamburg. InvD1640 was prepared in a buffer consisting of 20 mM HEPES pH 7.4, 200 mM NaCl and 5 mM DTT at a protein concentration of 3.5 mg/ml.

For InvE2448, SAXS data were measured at BM29 of the ESRF in Grenoble, France. InvE2448 was prepared in a buffer consisting of 20 mM HEPES pH 7.4, 200 mM NaCl and 5 mM DTT at a protein concentration of 1 mg/ml.

For RovC, SAXS data were measured on the P12 beamline at PETRAIII EMBL Hamburg. RovC was prepared in a buffer consisting of 50 mM TRIS pH 8.0, 500 mM NaCl, 5 mM DTT and 5% glycerol at a protein concentration of 2.5 mg/ml.

4 Results

4.1 InvD from YPIII strain of *Y. pseudotuberculosis*

Y. pseudotuberculosis YPIII possesses four different Invasins (InvA, InvB, InvC and InvD) on its surface. Unlike InvB and InvC, whose C-terminal domain share high sequence similarity with InvA, InvD C-terminal domain shares no sequence homology to any other Invasins. Thus, the aim of this part of work was to structurally and functionally characterize the recently identified InvD.

4.1.1 Cloning, expression and purification of InvD

4.1.1.1 Bioinformatics analysis of InvD

Sequence analysis of InvD using a combination of BLAST-p, Pfam (ID: B1JSC0/A0A0H3B1G5) and Signal-BLAST identified a putative signal peptide (SP in Figure 16) between residues 32 and 57. This N-terminal sequence stretch is followed by a β-barrel domain comprising of residues 146-421, which is required for anchoring the protein to the outer membrane of bacteria. The putative β-barrel domain is followed by 13 bacterial Ig-like (Big) domains. Big1-13 are capped by a predicted C-terminal domain (1829-1976 aa) with no similarity to known proteins. Based on these predictions, three constructs, InvD1640 (G1640-N1976), InvD1737 (P1737-N1976) and InvD1838 (P1838-N1976) were designed (Figure 16).

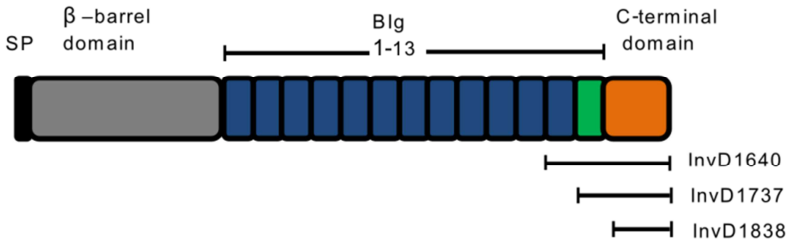

Figure 16: Domain organization of *Y. pseudotuberculosis* InvD. Scheme of the domain organization of *Y. pseudotuberculosis* InvD with construct boundaries.

4.1.1.2 Cloning and expression

InvD constructs were amplified by PCR from the genomic DNA of *Y. pseudotuberculosis* YPIII and cloned (3.2.1-3.2.7) into a modified pCOLA-Duet vector containing an N-terminal 6xHis tag (vector pVP009) and TEV cleavage site**.** After verification of inserts by sequencing (3.2.8), bacteria were transformed with the plasmid and tested for expression (3.3.1). Out of three constructs, InvD1640 and InvD1737 resulted in soluble protein, while InvD1838 was not soluble (Figure 17).

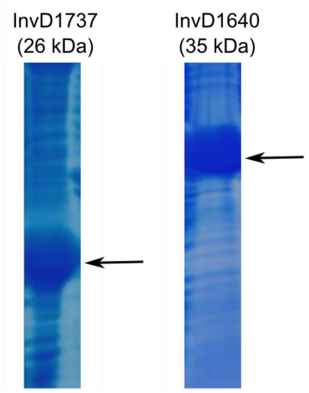

Figure 17: Coomassie stained 15% SDS gel of test expression of 6xHis InvD constructs. For both constructs, high yields of soluble protein were obtained by overnight incubation of cells at 20°C with 100 µM IPTG. Bands corresponding to soluble proteins are marked with the black arrow.

4.1.1.3 Culturing and cell disruption

As described in section 3.3.2-3.3.3, 4 L of TB + kanamycin + chloramphenicol media was inoculated with the overnight culture of InvD1640 or InvD1737 expressing bacteria. At an OD_{600} (optical density at 600 nm) of 0.7, expression was induced with 100 µM IPTG and cultures were allowed to grow at 20°C, 130 rpm for 21 h. After centrifugation and resuspension in lysis buffer, cells were homogenized.

4.1.1.4 Purification of 6xHis InvD constructs for crystallization

After clearing of the lysate by centrifugation, affinity chromatography was performed (3.3.4). To further purify proteins to homogeneity, size exclusion chromatography was performed (3.3.7) in 20 mM HEPES pH 7.4, 200 mM NaCl and 5 mM DTT. For InvD1640, S200 16/60 column was used, while InvD1737 was purified on S75 16/60 column (Figure 18).

Purified proteins were identified as *Y. pseudotuberculosis* YPIII InvD by tryptic digest and mass spectrometry. Proteins were concentrated to 40 mg/ml and flash frozen in liquid nitrogen. The overall yield of pure InvD1640 and InvD1737 was about 10 mg/L culture.

4.1.1.5 Protease stability assay

To assess the stability and flexibility of InvD in solution, InvD1640 and InvD1737 were treated with 12 different proteases.

Apart from treatment with Proteinase-K or Subtilisin, InvD1737 was highly stable with no major degradation detectable, and InvD1640 also showed high stability with only minor degradation (Figure 19).

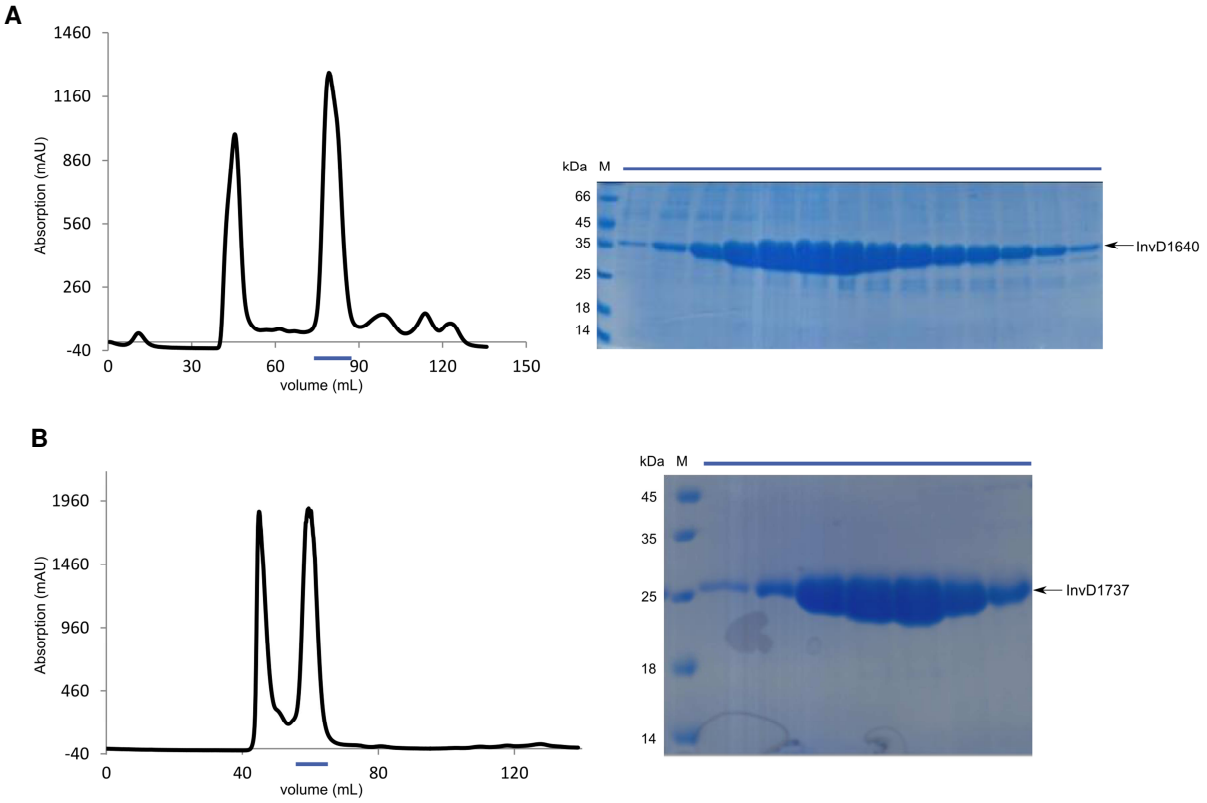

Figure 18: Purification of 6xHis tagged InvD constructs. A) Purification of InvD1640 and B) purification of InvD1737. Elution profile of the run on size exclusion column (left); Coomassie stained SDS gel of eluted fractions (right). Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS-PAGE. Absorbance of the protein is measured at 280 nm.

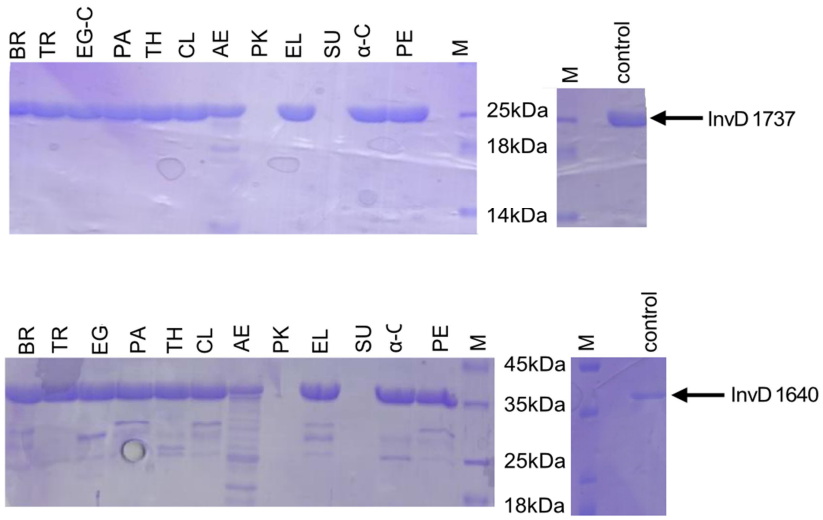

Figure 19: Stability of InvD against proteases. InvD1737 (top) and InvD1640 (bottom) (at a concentration of 1 mg/ml) were treated with 12 different proteases (BR-Bromelain, TR-Trypsin, EG-C-EndoproteinaseGlu-C, PA-Papain, TH-Thermolysin, CL-Clostripain, AE-Actinase, PK-Proteinase-K, EL-Elastase, SU-Subtilisin, α-C-α-Chymotrypsin , PE-Pepsin at a concentration of 0.01 mg/ml) and incubated for 20 h at room temperature.

4.1.2 Crystallization and structure solution of InvD

4.1.2.1 Crystallization of InvD

To gain insight into the structure of the adhesion domain of InvD, 6xHis tagged InvD1640 and InvD1737 were used for initial crystallization screening at 20°C and 4°C using commercially available screens. Initial hit with 22 mg/ml of InvD1640 was obtained in the following condition- 2 M ammonium dihydrogen phosphate and 0.1 M TRIS-HCl pH 8.5 at 20°C (Figure 20A). For optimization, crystals were grown using 24 well hanging drop plate at 20°C by mixing 0.5 µL of concentrated protein with 0.5 µL of mother liquor. Well diffracting quality crystals were obtained under the following condition-1.9 M ammonium dihydrogen phosphate and 0.1 M TRIS-HCl pH 8.0 with 18 mg/ml of InvD1640 (Figure 20B).

Several of these crystals were dissolved in protein buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 5 mM DTT) after washing in the reservoir solution and loaded on SDS-PAGE to confirm the identity of crystals. As a control, InvD1640 (in solution) was also loaded on the gel. The protein band was sent for mass spectrometry analysis which confirmed that the band in the gel belongs to InvD1640 (Figure 20C).

For cryoprotection, the reservoir solution was supplemented with an additional 25% PEG 400. Crystals were incubated for about 30 s in the cryosolution to ensure exchange of the reservoir solution and flash frozen in liquid nitrogen. InvD1640 crystals diffracted to a resolution of 3.6 Å in house (λ-1.5418 Å).

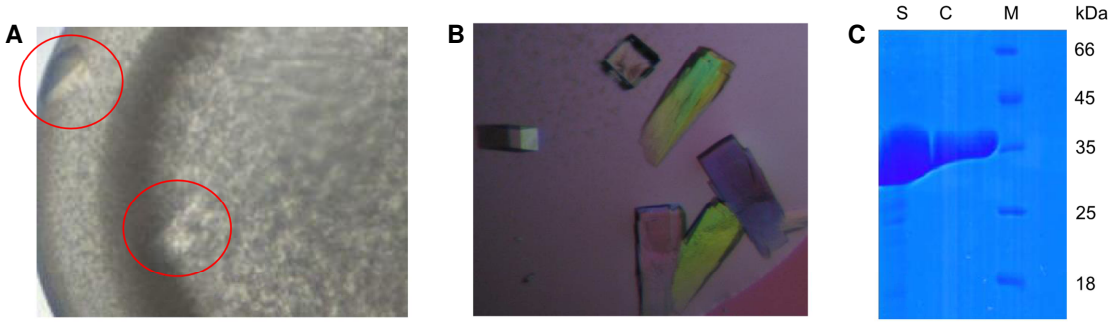

Figure 20: 6xHis tagged InvD1640 crystals. A) Initial hit (crystals are marked in red circle) B) optimized crystals and C) SDS-PAGE analysis of crystals (S=protein in solution, C=protein in crystals).

4.1.2.2 Data processing and structure solution

The native data set was collected at the Swiss Light Source (SLS) at X06DA/PXIII beamline. InvD crystals diffracted to 2.6 Å resolution. InvD C-terminal domain shows no significant similarity to other sequences in the protein database. Instead, BIg domains 12 and 13 share 31-53% sequence identity with BIg domains D1, D2 and D3 of InvA (PDB entry 1CWV (Hamburger, 1999)). Hence, initial molecular replacement trials were performed with Phaser MR by using individual Ig-like domains of InvA as a search template. A partial solution containing two Ig-like domains in the asymmetric unit was found. Further attempts to use the integrin binding domain of InvA (1CWV) as a search model did not improve the electron density.

Therefore to obtain the additional information on phases, heavy-atom derivatives were prepared by soaking the crystals in 1.9 M ammonium dihydrogen phosphate, 0.1 M TRIS-HCl pH 8.0 and 25% PEG400 supplemented with 100 mM of Yb-HPDO3A for 1 min, followed by flash freezing in liquid nitrogen.

The anomalous data set (1.38315 Å) was collected at the SLS at X06DA/PXIII beamline. The heavy atom soaked crystals diffracted to 3.3 Å. Phaser-EP was employed to calculate the phases. SAD alone failed to solve the structure. So, MR-SAD protocol was used in Phaser using the "SAD plus MR" mode. The model obtained after MR-SAD was subsequently used as the template to solve the native structure by molecular replacement.

InvD1640 crystallized in space group $p4_3$ with two molecules in the asymmetric unit (AU). Phenix-Autobuild was used for automated model building. Coot was used for further manual model building. The model was refined to 2.6 Å resolution with phenix.refine taking into account the twin law (h,-k,-l). Final step of refinement was performed using TLS refinement with the TLS groups as were defined by Phenix. Statistics of data collection and refinement are depicted in Table 24.The quality of the electron density is shown in Figure 21.

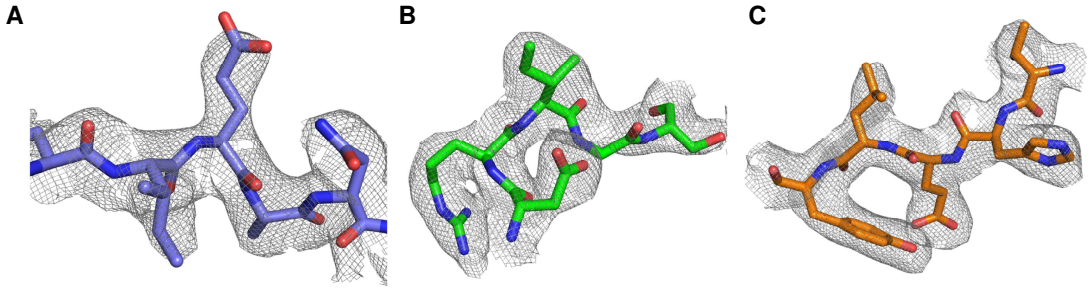

Figure 21: Quality of the electron density. $2F_o$-F_c electron density (contoured at 1 σ) in the A) BIg12 B) BIg13 and C) the C-terminal domain of InvD.

Table 24: Data collection and refinement statistics.

DATA COLLECTION	**InvD (PDB: 5LDY)**	**InvD (Yb-derivative)**
X-ray source	PXIII, SLS	PXIII, SLS
Wavelength [Å]	1.000	1.38315
Resolution [Å]	44.70-2.60 (2.70-2.60)	49.40-3.10 (3.20-3.10)
Space group	$P4_3$	$P4_3$
Cell dimensions		
a, b, c [Å]	58.9, 58.9, 274.5	58.7, 58.7, 274.0
α, β, γ [°]	90.0, 90.0, 90.0	90.0, 90.0, 90.0
R_{merge}**[%]**	6.3 (41.6)	8.0 (38.8)
R_{meas}**[%]**	6.6 (43.6)	8.3 (40.4)
CC1/2 [%]	99.9 (95.7)	99.9 (97.5)
‹I/σ(I)›	26.6 (5.8)	29.8 (6.8)
Completeness [%]	99.9 (99.9)	99.8 (99.5)
Unique reflections	28540 (3070)	33168 (3019)
Redundancy	11.3 (11.2)	12.4 (12.7)
REFINEMENT		
Resolution [Å]	44.70-2.60	
No. of reflections	28540	
R_{work}	0.230	
R_{free}	0.253	
No. of atoms	9725	
Avg. ***B*** **factor [Å^2]**	67.0	
R.m.s. deviations		
Bond length [Å]	0.0024	
Bond angles [°]	0.470	
Ramachandran [%] (favored/allowed/ disallowed)	90.6/9.4/0.0	

4.1.3 Structural characterization of InvD

4.1.3.1 Structure of InvD1640

The overall structure of the construct is rod-like, with a length of 12.9 nm as shown in Figure 22. InvD1640 consists of three domains: Big12 (1640-1736 aa), Big13 (1737-1837 aa) and the C-terminal domain (1838-1976 aa) which is named adhesion domain (AD) hereafter.

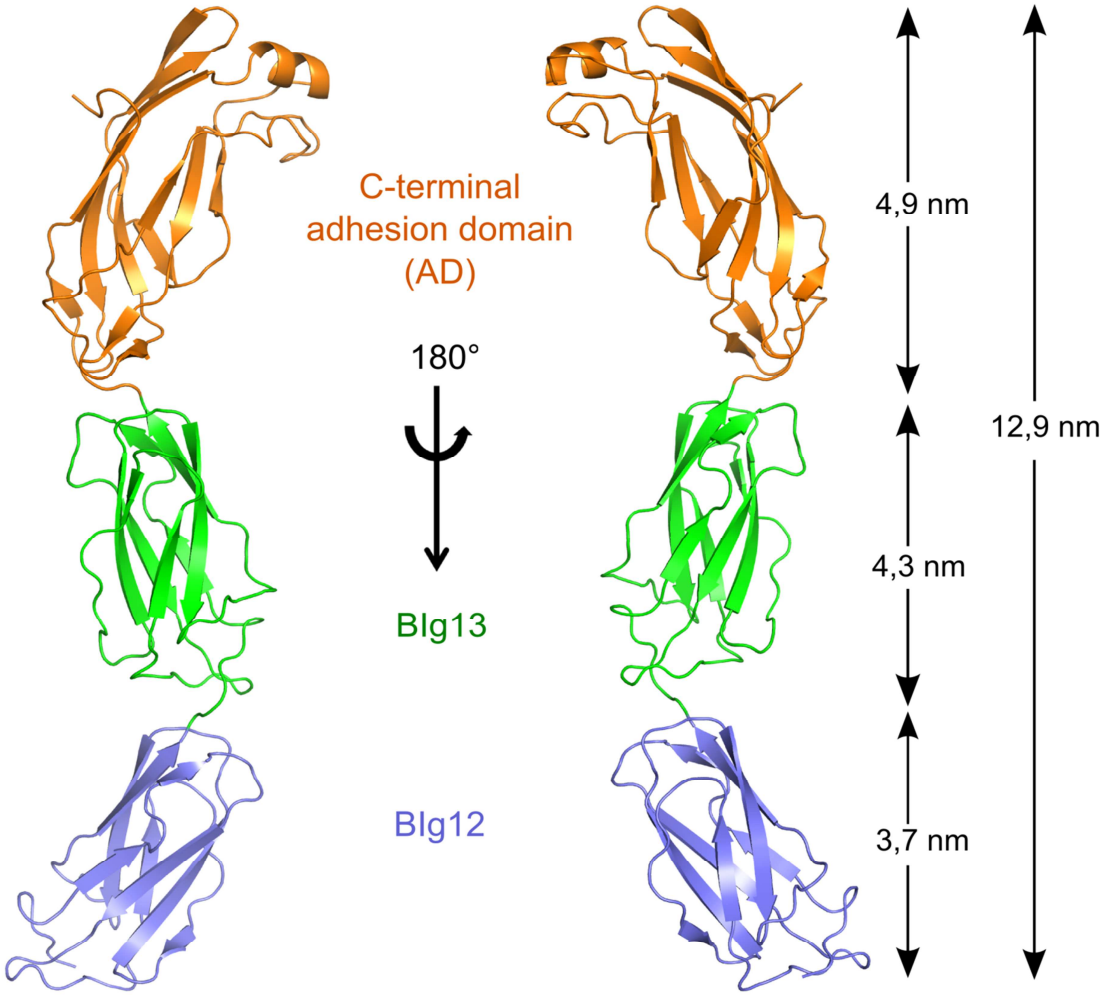

Figure 22: Crystal structure of *Y. pseudotuberculosis* InvD. Cartoon model of the crystal structure of InvD1640. InvD1640 consists of the N-terminal Big12 (blue), Big13 (green) and the C-terminal adhesion domain (AD in orange) shown in two orientations.

The two molecules in the asymmetric unit (AU) are very similar and they superimpose with an rmsd of 0.923 Å. The separate superimposition of Big13/AD or Big12 resulted in a lower rmsd of 0.814 Å and 0.761 Å, respectively.

4.1.3.2 SAXS structure of InvD1640

With the aim of understanding the structure of InvD in solution, SAXS data for InvD1640 was collected at the P12 beamline at PETRAIII EMBL Hamburg. The calculated SAXS envelope is highly similar to the crystal structure of InvD1640 although with the N-terminal BIg12 domain in a different conformation. Manual fitting of InvD1640 crystal structure to the envelope was performed by fitting separately the region corresponding to BIg13/AD and the N-terminal BIg12 (Figure 23A).

These small differences that were observed in the orientation of BIg12 suggest that the junction between BIg12 and BIg13 allows for minor flexibility.

The chi value of 1.09 was obtained with CRYSOL on fitting the SAXS data with the crystal structure, which indicates that the model is in very good agreement with the calculated SAXS-envelope (Figure 23B).

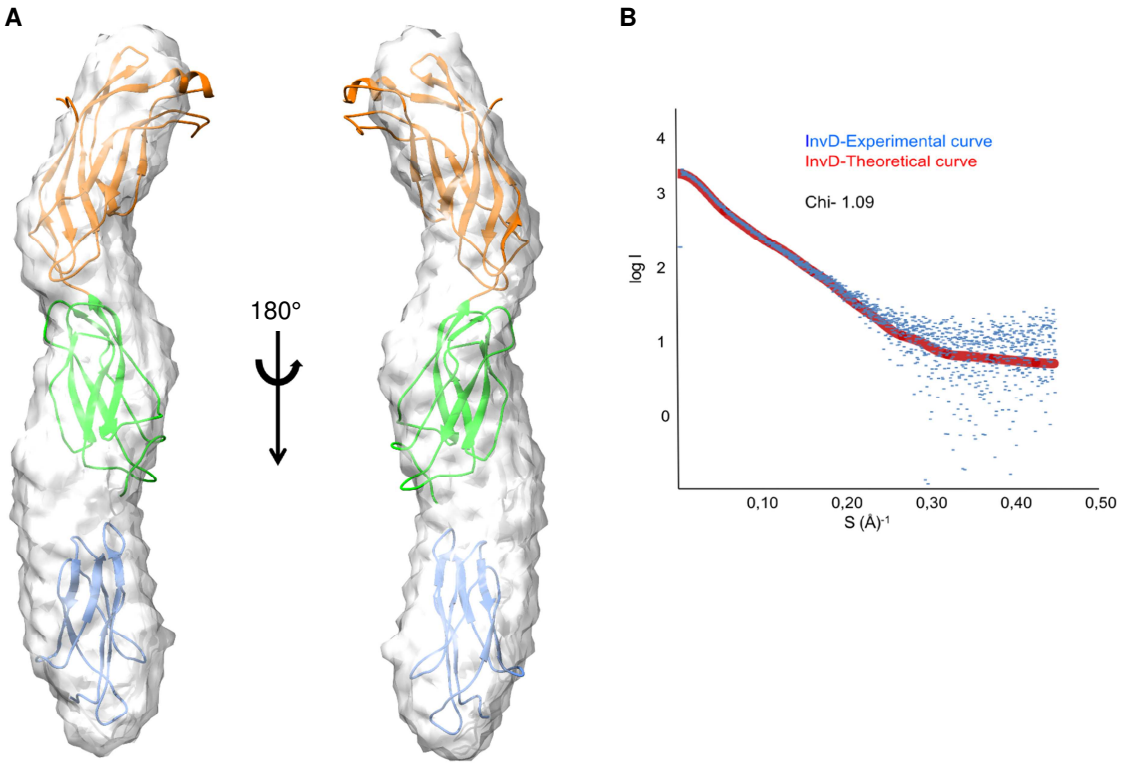

Figure 23: SAXS envelope of InvD1640. A) Rigid body fitting of the crystal structure into the *ab initio* determined SAXS envelope shown in two orientations. B) Fit of the rigid body model (red) with the experimental scattering (blue). Quality of the fit is expressed in terms of chi value.

4.1.3.3 Topology of InvD

Big12/13 domain

InvD comprises 13 Big like domains; of these, the first 12 Big-domains share high sequence identity in the range of 56-97% (72-100% similarity), whereas Big13 is significantly different from Big1-12 (sequence identity 22-28%; similarity 40-47%) (Figure 24).

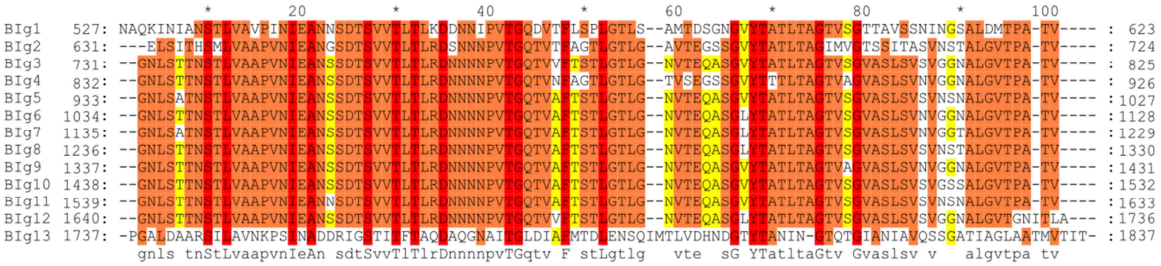

Figure 24: Sequence alignment of Big 1-13 domains of InvD. 100% conserved amino acids are marked in red, 80% in orange and 60% in yellow. Alignment has been generated with GeneDoc (Nicholas et al., 1996).

Despite the low sequence identity between Big12 and Big13, their structures are highly similar, each adopting a classical Immunoglobulin Superfamily (IgSF) fold (Figure 26 and Figure 27).

IgSF fold consists of two β-sheets with seven or more strands adopting a greek-key topology. IgSF domains have been previously classified into three different sets- C, V and I-set (Figure 25). The core structure of these domains is similar; variation arises due to the presence of different number of strands and their arrangement. The C and V domains are so named due to the similarities which they share with the constant and variable domains of antibodies.

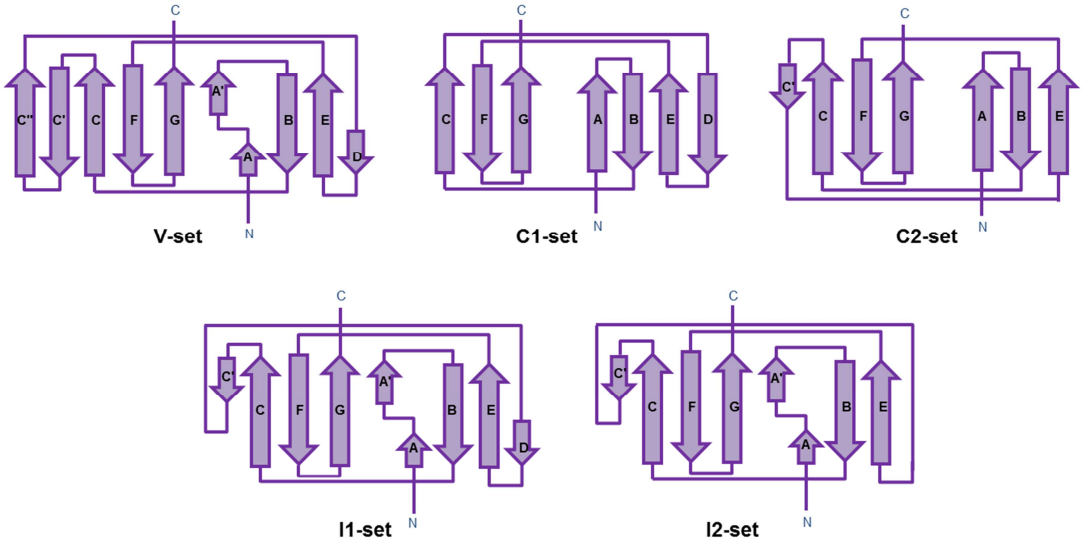

Figure 25: Topology diagrams of five different sets of the immunoglobulin fold. Variable (V-set), constant1 (C1-set), constant2 (C2-set), intermediate1 (I1-set) and intermediate2 (I2-set) (modified from (Bodelón et al., 2013)).

The C set is further divided into C1 and C2-sets. Structures grouped in C1-set consist of strands ABED formed by one β-sheet while the second sheet comprises β-strands G, F and C. Structures that belong to C2-set lack the D strand, but contain a C' strand. Structures that belong to V set comprise strands ABED forming one β-sheet, whereas the second sheet consists of A'GFCC'C" strands. I-set is intermediate in topology between C and V sets. In I-set, one β-sheet is composed of the ABED strands while the other sheet comprises A'GFCC' strands. Like C-set, I-set is also classified into I1 and I2 sets. Structures that belong to I1-category have a D strand while members of the I2 do not (Figure 25) (Bodelón et al., 2013; Casasnovas et al., 1998).

In Big12 and Big13 domains of InvD, one β-sheet is composed of strands ABED, while the second β-sheet is composed of strands A'GFC (Figure 26). Both Big12 and Big13 have a kink in strand A-A' mediated by cis-Proline (P1653). These results, together with the presence of strand D as well as the lack of strand C" are reminiscent of the I1-set of IgSF domains (Figure 25 and Figure 26). Even though Big12 and Big13 lack strand C', both Big12 and Big13 were classified in I1-set, in line with the previously observed lack of C' in other I1-set domains (Tan et al., 1998).

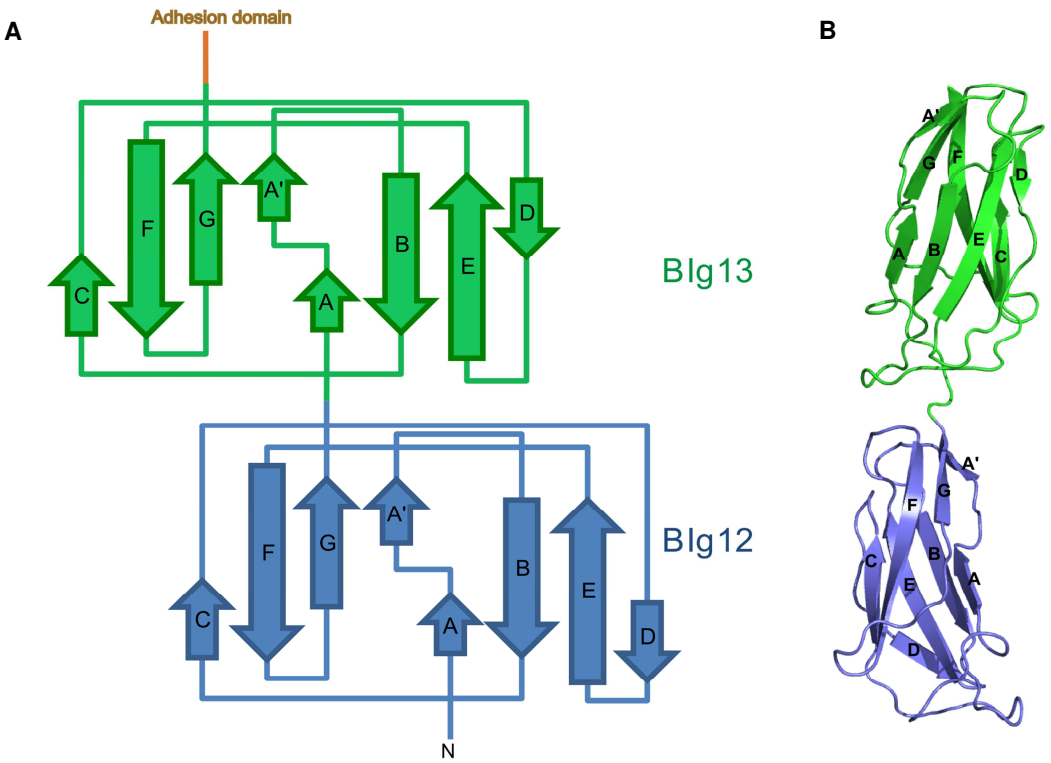

Figure 26: Domain architecture of Big12/13 domains. A) Topology diagram of Big12 (blue) and Big13 (green) of InvD1640**.** (B) Cartoon representation of the Big12/13 domains of InvD (β-strands are labeled).

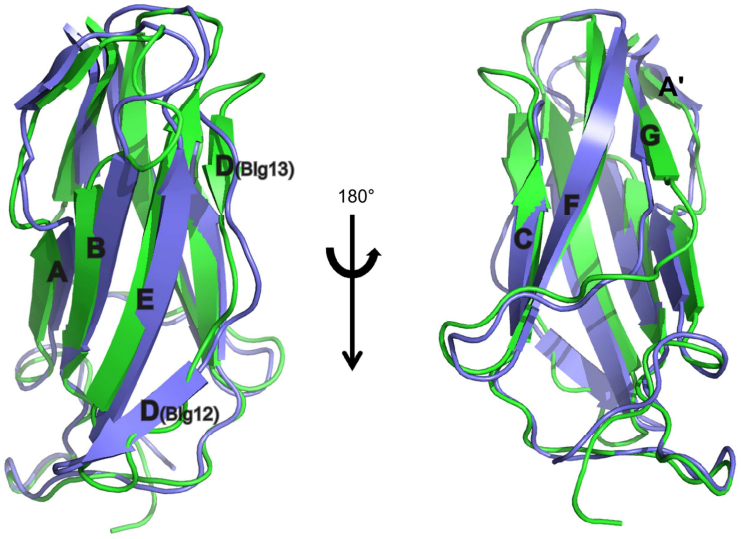

Figure 27: **Structural overlay of Blg12 and Blg13.** Superposition of Blg12 (blue) and Blg13 (green) shown in two orientations (β-strands are labeled). They superimpose with an rmsd of 1.58 Å.

Both Blg12 and Blg13 lack the usual disulfide bridge that connects the two sheets of the β-sandwich fold as well as the core residues conserved in IgSF structures, such as the tyrosine-corner or the C-strand tryptophan. The canonical cysteines in strands B/F are replaced by an interacting set of hydrophobic residues comprising V1666/L1718 and I1764/I1817 for Blg12 and Blg13, respectively (Figure 28) (Jones et al, 1992).

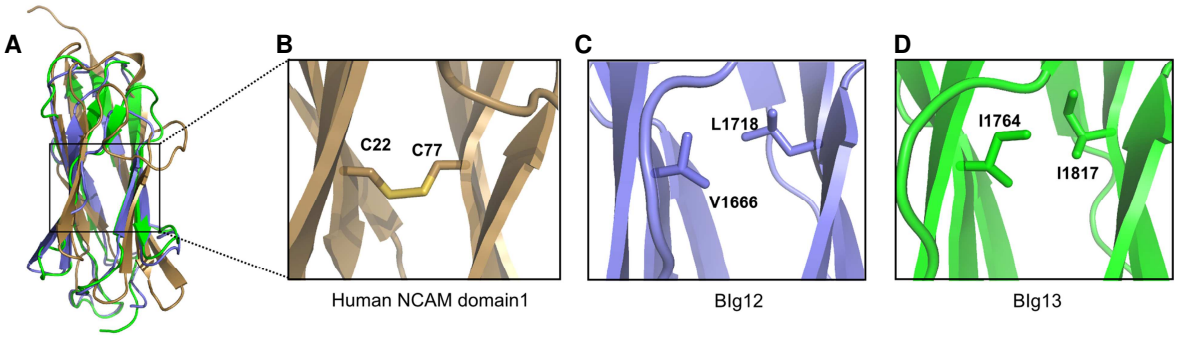

Figure 28: Close-up view of the hydrophobic core of Blg12/13 and its comparison with human NCAM domain-1. A) Structural overlay of Blg12 (blue), Blg13 (green) and Human NCAM domain1 (brown, PDB code 5AEA). B) Human NCAM domain-1 adopts a classsical Ig-like fold (Griffiths et al., 2016) and contains a disulfide bond formed by C22 and C77 while in C) Blg12 and D) Blg13 cysteines were replaced by hydrophobic amino acids. These residues are shown as sticks with sulfur atoms in yellow.

Blg12/13 interface

The contact area between Blg12 and Blg13 comprises an interface area of 199 $Å^2$ comprising a total of 13 interacting residues as determined by the PISA server (Krissinel, 2015) (Figure 29A, Table 25).

Table 25: PISA predicted Blg12-13 and Blg13-AD interacting interface.

	Blg12-Blg13 interface	**Blg13-AD interface**
N_{res}	13 (6 on Blg12 and 7 on Blg13)	19 (9 on Blg13 and 10 on AD)
N_{at}	47	69
Interface area (Å^2)	199.0	317.1
N_{HB}	2	4
N_{SB}	0	0
ΔG (kcal/mol)	-1.4	-3.9

N_{res}: indicates the number of interfacing residues; N_{at}: indicates the number of interfacing atoms; Interface area: half sum of the buried surface area; N_{HB}: number of hydrogen bonds in the interface N_{SB}: number of salt bridges in the interface; ΔG: gain in solvation energy.

Apart from contributions from the Blg12-Blg13 linker region (G-A), interface residues on Blg12 are located on the interstrand loops A'-B and E-F, whereas on Blg13 interface residues cluster exclusively on the B-C interstrand loop (Figure 29A).

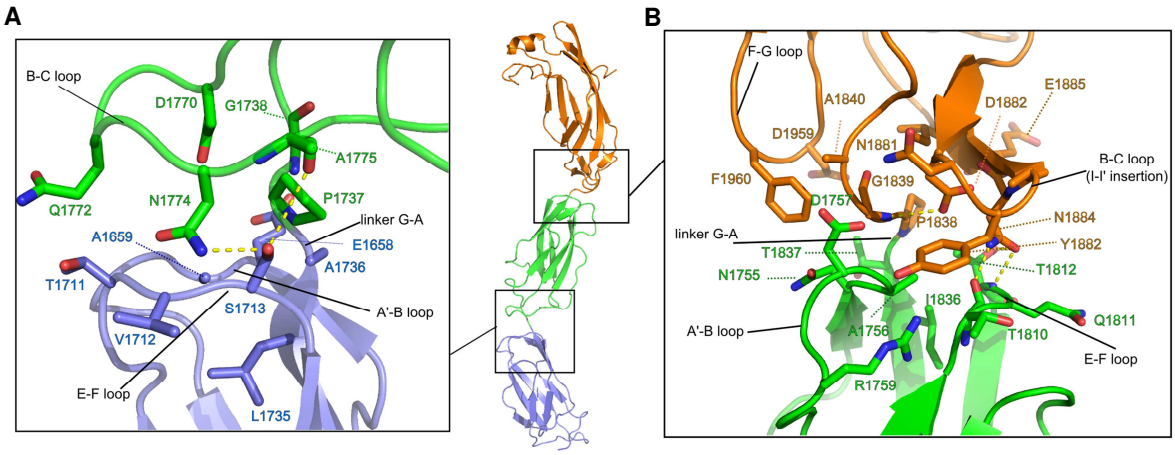

Figure 29: Blg12/13 and Blg13/AD domain interface. Detailed description of the inter-domain junctions of the A) Blg12–Blg13 and B) Blg13–AD. Residues responsible for forming interface contacts are labeled and shown as sticks with oxygen atoms in red and nitrogen atoms in blue. Hydrogen bonds are shown as yellow dashed lines. Interstrand loops are labeled.

While these interactions stabilize the relative orientation of Blg12-Blg13 and rod-like arrangement, flexibility of Blg12 in solution, as seen in the SAXS data (4.1.3.2), can be explained by this small interface area.

Adhesion domain of InvD

The C-terminal adhesion domain comprises residues 1838-1976 with a molecular weight of 14.8 kDa. As illustrated in Figure 30A and B, the structure of the AD is composed of β-

strands with a single α-helix at the tip of the rod shaped protein. The AD shares 37% and 27% sequence identity to BIg12 and BIg13, respectively. This is also reflected in the high degree of structural similarity of the two BIg domains with the core structure of the adhesion domain (Figure 30C). InvD AD core is formed by two antiparallel β-sheets reminiscent of the IgSF, in which ABE strands form one β-sheet, while the second β-sheet is formed by four antiparallel GFCC' strands. The lack of strand A/A' subdivision, lack of strand D combined with the presence of strand C' are reminiscent of the C2-set of IgSF.

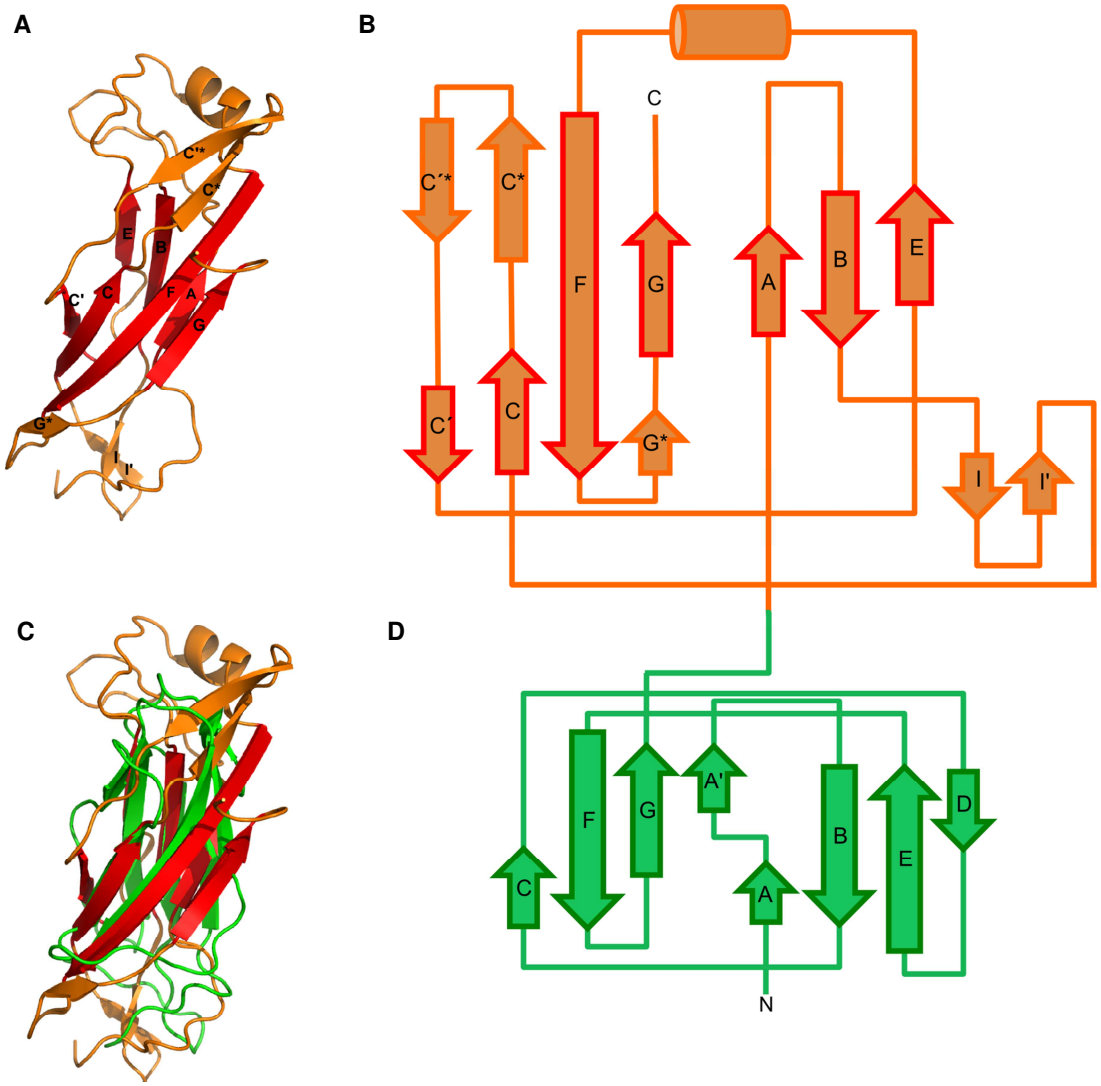

Figure 30: Topology of the adhesion domain and its comparison with BIg21. A) Cartoon representation of the AD of InvD (β-strands are labeled). B) Topology diagram of the AD. C) Structural superposition of BIg13 and the AD. They superimpose with an rmsd of 2.5 Å. D) Topology diagram of BIg13. Secondary structure elements that are unique to the IgSF fold in the AD are marked in red in (A), (B) and (C).

However, various insertions and modifications are observed in the structure of InvD AD apart from the canonical IgSF fold: i) an elongated linker region is present between strand A and B; ii) a short helix is inserted between strands E and F, that also includes a disulfide bridge

formed between C1859 and C1939; iii) two short antiparallel β-strands (I/I'), that are inserted between strands B and C and enlarge the interaction surface with Big13. In addition, iv) a further insertion between strands C and C' composed of two antiparallel β-strands (C'*/C*) is present, which in combination with v) the extended strand F form a third antiparallel three stranded β-sheet (C'*C*F), sharing strand F with the canonical GFCC' sheet (Figure 30A).

Big13/AD interface

Compared to Big12-Big13, the Big13-AD interdomain interface spans a significantly larger area of 317 Å^2 attributed to 19 interacting residues in total (Figure 29B, Table 25). Apart from contributions from the Big13-AD linker region (G-A), on Big13, these interactions are clustered on interstrand loop A'-B and E-F, whereas on the AD, residues are clustered in the short antiparallel β-strand insertion between strands B and C and in the interstrand linker F-G (Figure 29B and Figure 30B).

4.1.3.5 Structural similarity with InvD-AD

Search for structural similarity with the AD using PDBefold server (http://www.ebi.ac.uk/msd-srv/ssm) revealed weak structural (Z-score < 3.2) but no sequence similarity with variety of proteins of different functions such as adnectin, a monobody that is engineered to target specific proteins (Lipovšek, 2011), PsaA from *Y. pestis* which is shown to agglutinate red blood cells, binds to carbohydrates and lipids and also inhibits phagocytosis by macrophages by binding to the Fc part of IgG (Bao et al., 2013) and SafA pilin, a protein involved in biofilm formation (Remaut et al., 2006) (Figure 31).

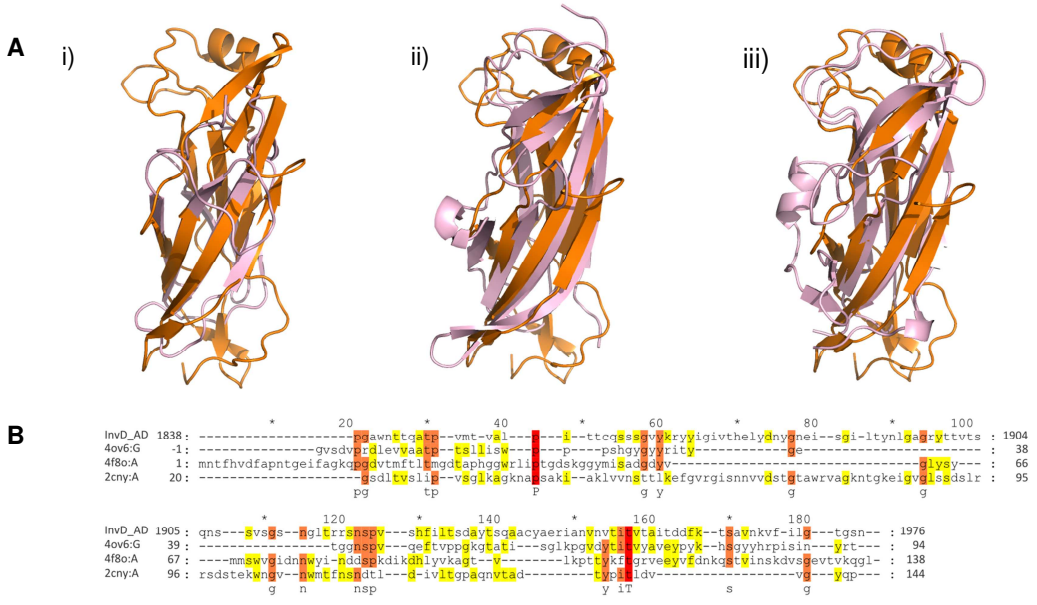

Figure 31: Comparison of InvD-AD with other structurally similar proteins. A) Structural Superposition of InvD-AD (orange) with (in pink): i) Adnectin (PDB code-4OV6:G) (rmsd 4.3 Å), ii) PsaA (PDB code-4F8O:A) (rmsd 3.7 Å) and iii) SafA pilin (PDB code-2CNY:A) (rmsd 3.7 Å), B) Sequence alignment of InvD-AD with Adnectin (4OV6:G), PsaA (4F8O:A) and SafA pilin (2CNY:A). (100% conserved amino acids are marked in red, 80% in orange and 60% in yellow). Alignment is generated with GeneDoc (Nicholas et al., 1996).

Despite the structural homology of the AD core to IgSF, taking the insertions into account, the AD adopts a distinct fold. These observations made it difficult to assign function of InvD AD on the basis of its structural features.

4.1.4 Functional characterization of InvD

Rebecca Geyer demonstrated that InvD is expressed during infection and required for efficient colonization of the host (Dissertation Rebecca Geyer, 2014). But the information on the identity of the host interaction partner to which InvD targets was not known.

In analogy to InvA, the AD in proteins of the Invasin-family usually represents the domain that targets a specific host cell surface molecule (Pisano et al., 2012; Strong et al., 2011). With the aim to identify the molecular target and to characterize the function of InvD, binding of recombinant InvD protein to carbohydrates, proteins and lipids was systematically tested.

4.1.4.1 Cloning, expression and purification of InvD and InvA

Cloning

For receptor identification, InvD (G1640-N1976) was cloned into a modified pCOLA-Duet vector containing N-terminal 6xHis, TEV-protease cleavage site and 3xflag tag (vector pPS18) or N-terminal strep affinity tag with TEV cleavage site (vector pVP008). As a negative control, InvD construct that lacks the AD, hereafter referred to as ΔAD-InvD (G1640-G1839) was designed and cloned into the above mentioned vectors.

It is known that InvA binds β_1- integrins on the surface of human epithelial cells (Isberg et al., 1987). Therefore, for the beads assay and MST experiments (please refer to 4.1.3.2 and 4.1.3.5), InvA was used as a control. InvA (P500-I985) was PCR amplified by using pET28 plasmid containing InvA (kindly provided by Prof. Petra Dersch) and cloned into a vector containing a N-terminal 6xHis_TEV_3xflag site.

Expression of InvA or InvD constructs

Strep-tagged InvD1640 or ΔAD-InvD, 6xHis_3xflag-tagged InvD1640 or ΔAD-InvD or InvA were expressed in 3 L *E. coli* culture by overnight incubation of cells at 20°C with 100 µM IPTG as described in 4.1.1.2.

Purification of strep tagged InvD1640

Strep-tagged InvD1640 and ΔAD-InvD were first purified by strep affinity. Elution fractions were pooled and loaded on size exclusion column to further purify the protein (Figure 32).

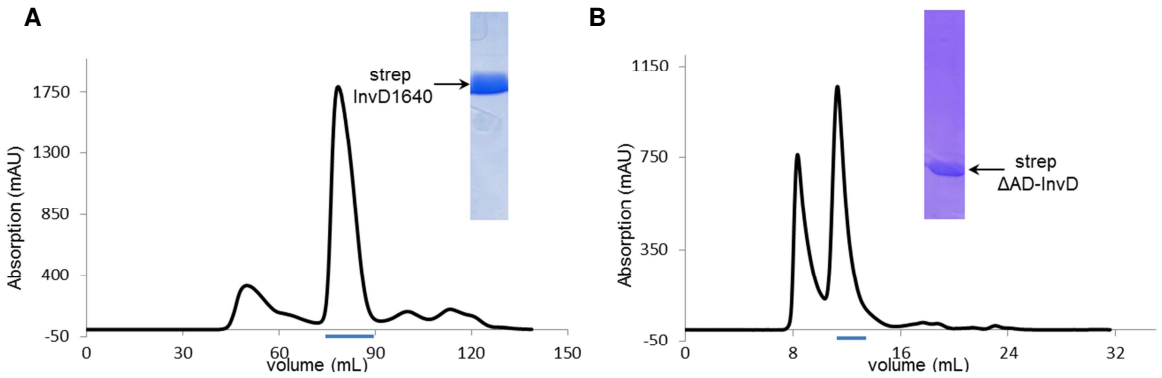

Figure 32: Purification of strep tagged InvD constructs. Elution profile of strep tagged A) InvD1640 on S200 16/60 column B) ΔAD-InvD on S75 10/30 column and corresponding SDS gel showing the purity of proteins. Absorbance of the protein is measured at 280 nm. The fractions indicated by the solid blue bar were pooled and concentrated.

Purification of 6xHis_3xflag tagged InvA and InvD constructs

6xHis_3x flag tagged constructs were first purified by Ni-affinity. Elution fractions were pooled and incubated with TEV to remove 6xHis tag. The 3xflag-tag was retained for later identification with a highly specific anti-Flag antibody. Cleaved proteins were loaded on size exclusion column to purify the protein to homogeneity (Figure 33 and Figure 34).

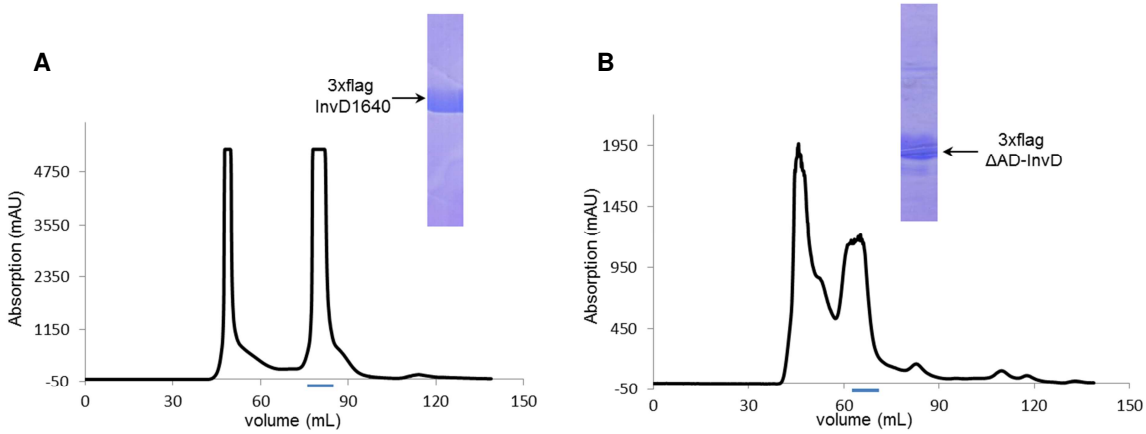

Figure 33: Purification of 3xflag tagged InvD constructs: Elution profile of 3xflag tagged A) InvD1640 on S200 16/60 column. B) ΔAD-InvD on S75 16/60 column and corresponding SDS gel showing the purity of proteins. Absorbance of the protein is measured at 280 nm. The fractions indicated by the solid blue bar were pooled and concentrated.

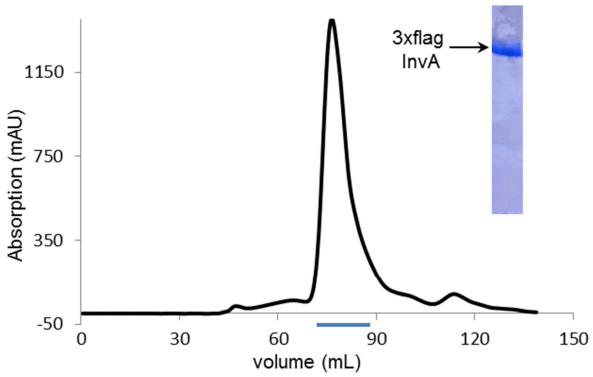

Figure 34: Purification of 6xHis_3xflag tagged InvA construct. Elution profile on S200 16/60 column and corresponding SDS gel showing the purity of the protein. Absorbance of the protein is measured at 280 nm. The fractions indicated by the solid blue bar were pooled and concentrated.

4.1.4.2 InvD attachment to Human epithelial (HEp-2) cells

To identify if InvD mediates attachment to HEp-2 cells, similar to InvA, the attachment efficiency of InvD1640 coated latex beads on HEp-2 cells was analyzed. Significant number of beads coated with InvA was found attached on the cell surface, while beads coated with InvD or BSA showed no efficient cell attachment (Figure 35A). So, the possibility that InvD targets a molecule, which is commonly exposed on the cellular surface of HEp-2 cells, was excluded.

4.1.4.3 InvD binding to lipids and carbohydrates

Since binding to a factor exposed on the surface of HEp-2 cells could be excluded, binding of 3xflag tagged InvD1640 to membrane lipid strip was tested. But no binding of InvD1640 was observed in a lipid binding assay (data not shown). Similarly, 3xflag tagged InvD1640 was screened with the mammalian glycan array (within the Consortium for Functional Glycomics (CFG)). Strong signal to glycans 540 (1000 Relative Fluorescence Units (RFU)), 544 (550 RFU) and 582 (264 RFU) were observed, but these signals were also observed with the negative control, showing that overall no binding to glycans was detected (table with raw data shown in appendix 8.2.1) (Figure 35B).

Since no glycan or lipid was identified as the interaction partner, as a next step, InvD binding with proteins in a cell extract was tested.

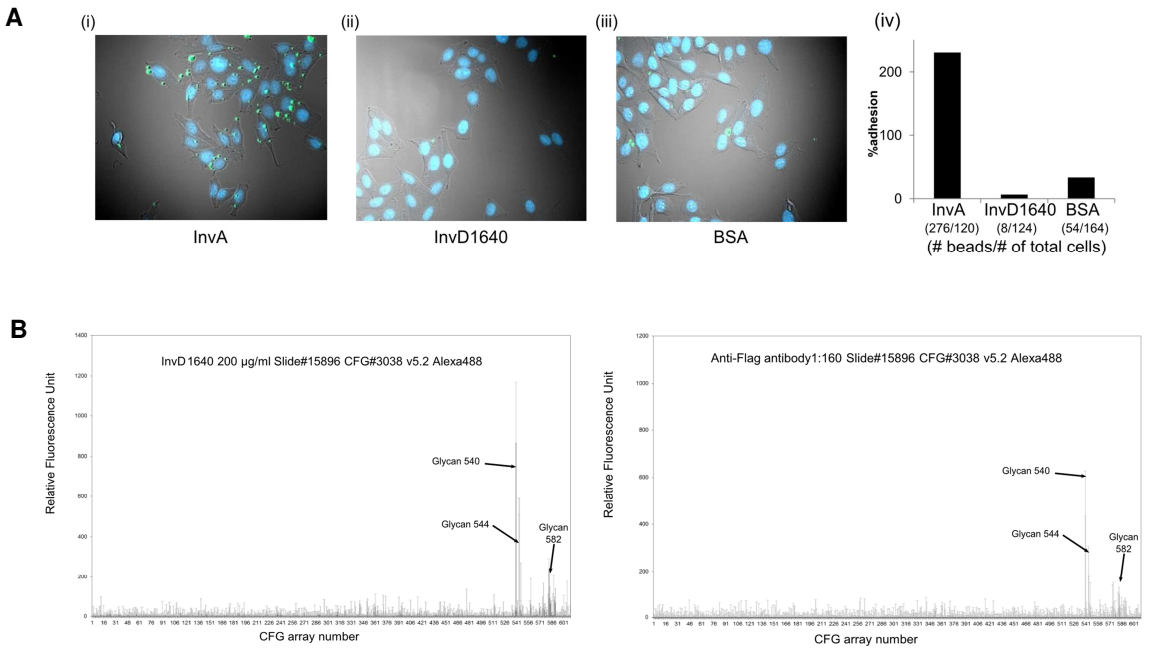

Figure 35: InvD1640 binding to HEp-2 cells, lipids and carbohydrates. A) Latex beads (in green) coated with (i) InvA (positive control), (ii) InvD1640 (iii) BSA (negative control) were incubated with HEp-2 cells. (iv) Attached beads per cell were quantified after removal of unattached beads. Numbers in brackets correspond to (# beads/ # of total cells). B) Glycan binding analysis of 3xflag tagged InvD (left); As a control, binding of anti-flag antibody on the array was also analyzed (right). Glycans for which strong signal were observed, are marked with the black arrow.

4.1.4.4 Pull down experiments with homogenized mice organs

To investigate if InvD1640 targets proteins as its receptor, pull down experiments using 3xflag tagged InvD1640 with homogenized extract of uninfected mice peyer´s patches and intestine (small intestine, caecum and colon) was performed (3.5.1) (Figure 36).

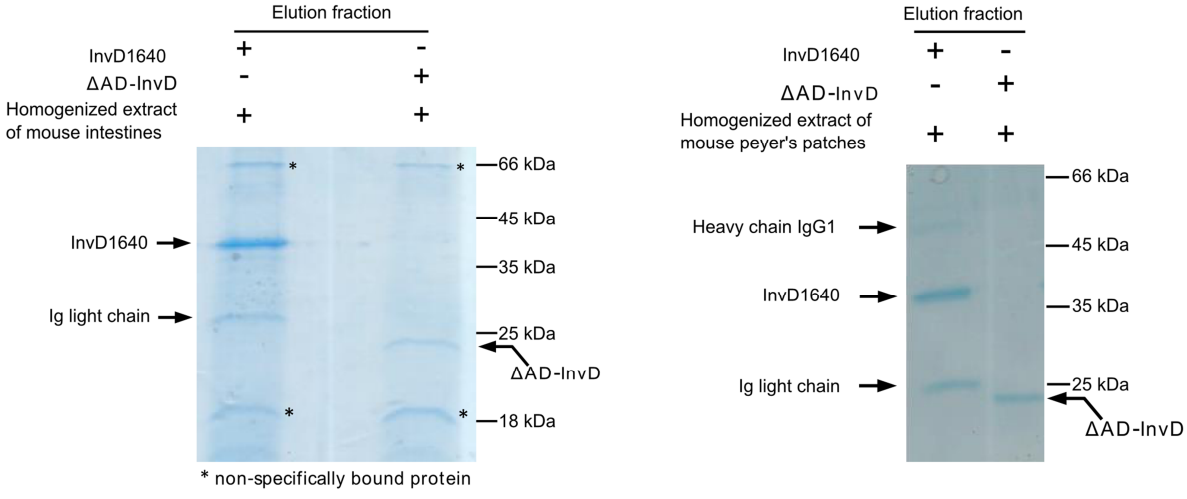

Figure 36: InvD interaction with immunoglobulins. Homogenized extract of mice intestines (left) and peyer's patches (right) mixed with 3xflag tagged InvD1640 or ΔAD-InvD were subjected to flag pulldown assay.

As a negative control, 3xflag tagged ΔAD-InvD was used. Elution fractions obtained after pull downs were analyzed by SDS-PAGE. In InvD1640 elution fraction, a 25 kDa and a less prominent 50 kDa bands were observed, while these bands were absent in ΔAD-InvD elution fraction. The 25 kDa and 50 kDa proteins were identified as immunoglobulin light and heavy chain by mass spectrometry suggesting that InvD targets the immunoglobulins (Figure 36).

4.1.4.5 Microscale thermophoresis

To confirm the interaction between InvD and immunoglobulins, MST was used (3.5.7). Cy5-labelled InvD1640, ΔAD-InvD and InvA were titrated with increasing concentrations of mouse IgG, human IgG and human IgA (Figure 37A, B and C). The obtained results confirmed the binding of InvD with IgG/IgA with an experimentally determined affinity constant (K_d) in the range of 1 to 10 µM. No binding was observed when labeled controls ΔAD-InvD or InvA were titrated with human/mouse IgG or human IgA, confirming an AD-dependent specific interaction. In order to narrow down the region on immunoglobulins to which InvD binds to, binding of Cy5 labeled InvD1640 was tested with IgG Fc and IgG F(ab')$_2$. The results showed that InvD specifically targets the F(ab')$_2$ fragment (Figure 37D).

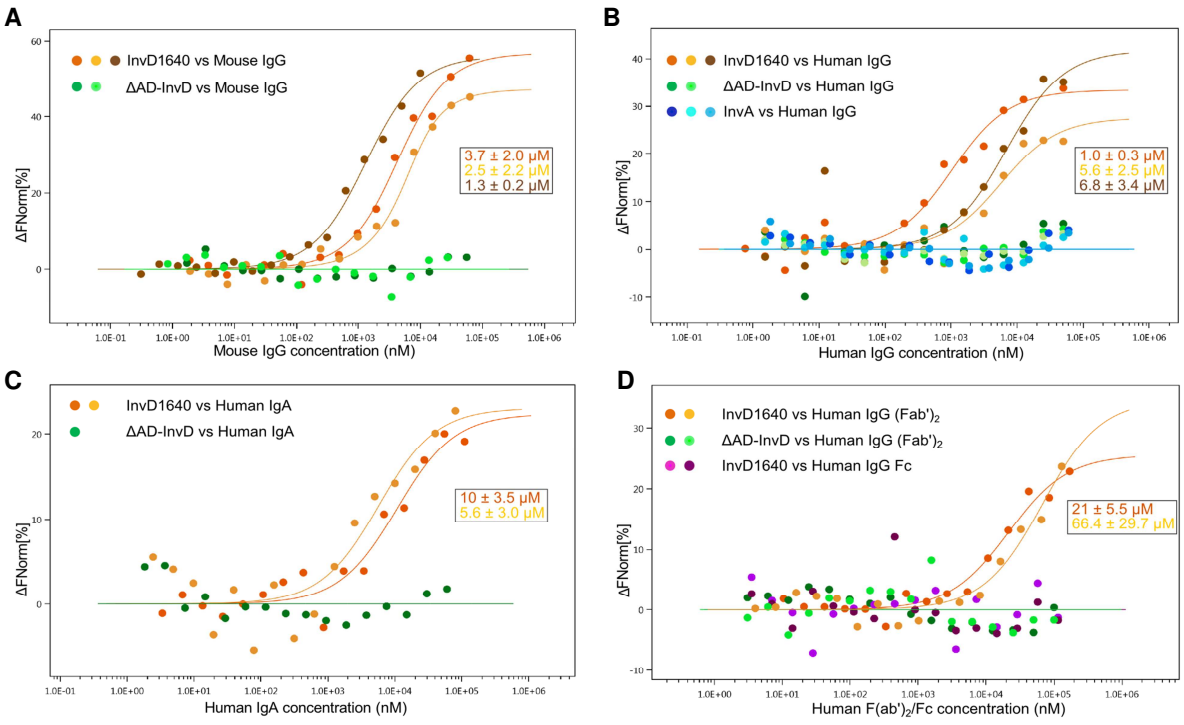

Figure 37: InvD interaction with immunoglobulins. Binding affinities of InvD1640 with mouse IgG (A), human IgG (B), human IgA (C) and human F(ab')$_2$ or Fc (D) were determined by Microscale thermophoresis. InvA or ΔAD-InvD were used as negative control. Individual measurements are shown performed with different concentrations of antibodies.

4.1.4.6 InvD specifically binds to B-cell populations isolated from mouse

Based on MST data (Figure 37D), it was clear that InvD targets the Fab fragment of the soluble antibody. Since the antibody structure is also present on B cells in the form of B cell receptor, a flow cytometry approach was used to analyze if InvD can target B-cell receptor present on B cells apart from targeting soluble antibodies.

Binding of 3xflag tagged InvD1640 to *ex vivo* isolated murine B-cell populations was tested via flow cytometry (3.5.4) in collaboration with the group of Prof. Jochen Hühn (experiments performed by Jörn Pezoldt). Only B cells were enriched for cells bound by InvD1640, whereas total T cells, γδT cells, monocytes, NK cells and dendritic cells did not show any binding of InvD1640 (Figure 38B), indicating that InvD binds specifically to the $F(ab')_2$ fragment of the B-cell receptor. Moreover, ΔAD-InvD showed a 10-fold lower signal in frequency and intensity, further validating that the binding of InvD1640 relied on the adhesion domain (Figure 38A).

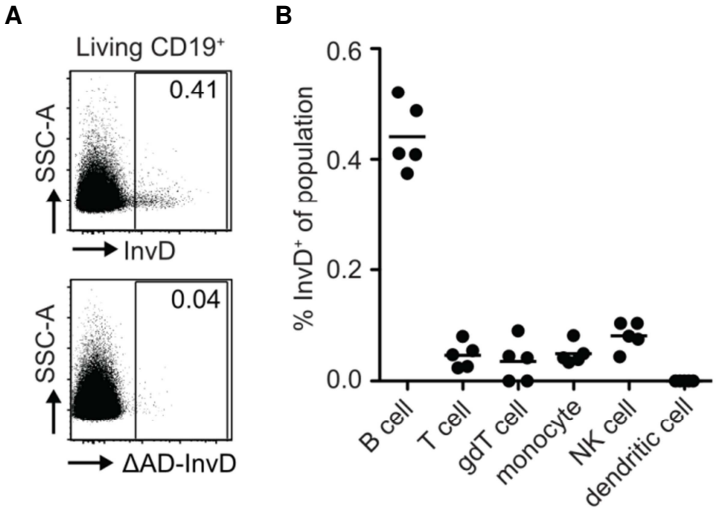

Figure 38: Single cell suspensions from indicated organs were analyzed using flow cytometry. A) Exemplary dotplots depict frequencies and fluorescence intensity of 3xflag tagged InvD1640 or 3xflag tagged ΔAD-InvD on living B cells using anti-flag-antibody conjugated to APC. B) Scatterplot summarizes frequencies of $InvD^+$ cells of the respective cell population. Pooled data of two independent experiments (n = 2-3). B cells: $CD19^+$, T cells: $CD3^+$: γδT cells: $CD19^-gdTCR^+$, NK cell: $CD49b^+$, monocytes: $CD19^-γδTCR^-CD11b^+CD11c^-$, dendritic cells: $CD19^-γδTCR^-CD11b^{med}CD11c^+$.

4.1.4.7 Characterization of Fab binding specificity by phage display panning

To gain a detailed insight into the specificity of interaction of InvD with Fab, phage display panning was performed (3.5.2) in collaboration with the group of Prof. Michael Hust (experiments performed by Saskia Helmsing). HAL10 (kappa library) and HAL9+10 (combination of kappa and lambda library) libraries were separately incubated with immobilized strep tagged InvD1640. After three rounds of panning, eluted phages were reamplified in *E. coli* and plated on agar plate. 96 clones were randomly picked and phage

ELISA was performed on immobilized strep tagged InvD1640 and ΔAD-InvD (Figure 39A, B). Compared with ΔAD-InvD, 20-60 fold higher absorbance signal was observed with InvD1640 indicating that the interaction of InvD with scFvs is AD specific.

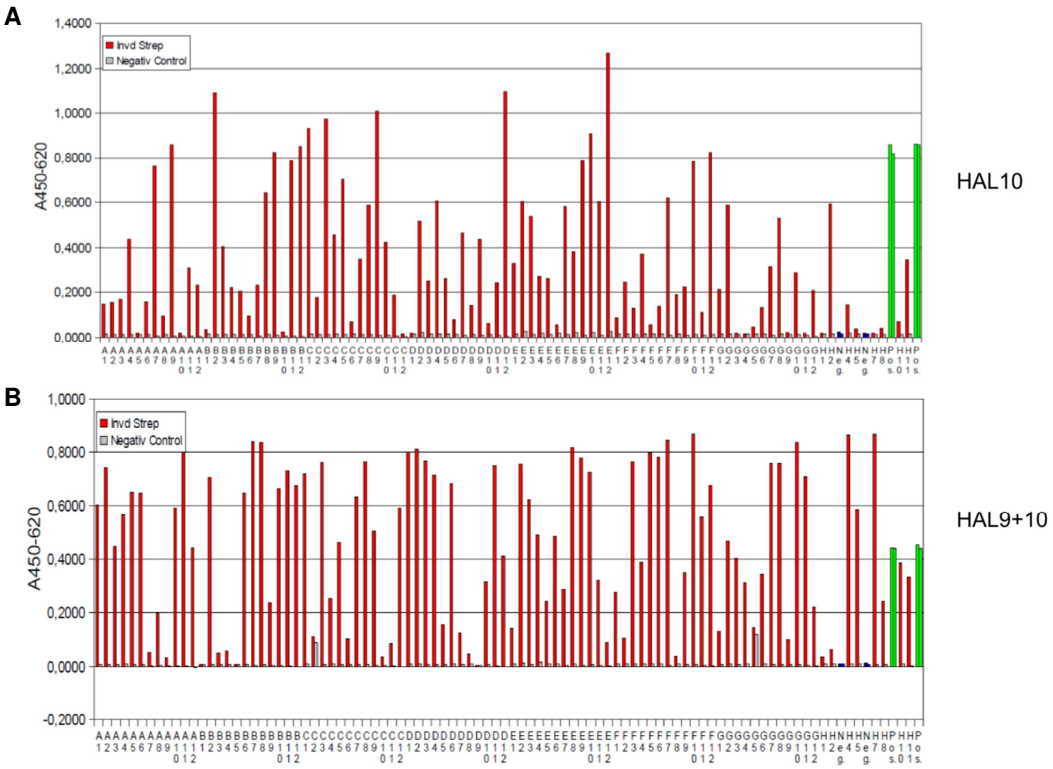

Figure 39: Binding assessment of soluble scFv to InvD1640 by ELISA. Binding of soluble scFv from A) HAL10 and B) HAL9+10 library to InvD1640 (Invd strep) and ΔAD-InvD(Neg. control) was analyzed by microtiter plate ELISA. Green bar represents the positive control (lysozyme + anti-lysozyme scFv), blue bar represents the background signal of the system (streptavidin + *E. coli* supernatant + components of detection system).

Phagemid DNA of 57 and 59 randomly selected individual scFvs after 3rd round of panning on InvD1640 with HAL10 and HAL9+HAL10 library, respectively, were sequenced. Sequence analysis of scFvs showed high specificity for VK1 subfamily in combination with VH3 subfamily. While this combination only represents approx. 16% and 6% of the unselected HAL10 and HAL9+10 library, 86% and 59% of the clones isolated after panning correspond to VH3/VK1, respectively (Figure 40A, B), giving rise to the notion that InvD specifically targets antibodies of the VH3/VK1 subfamily.

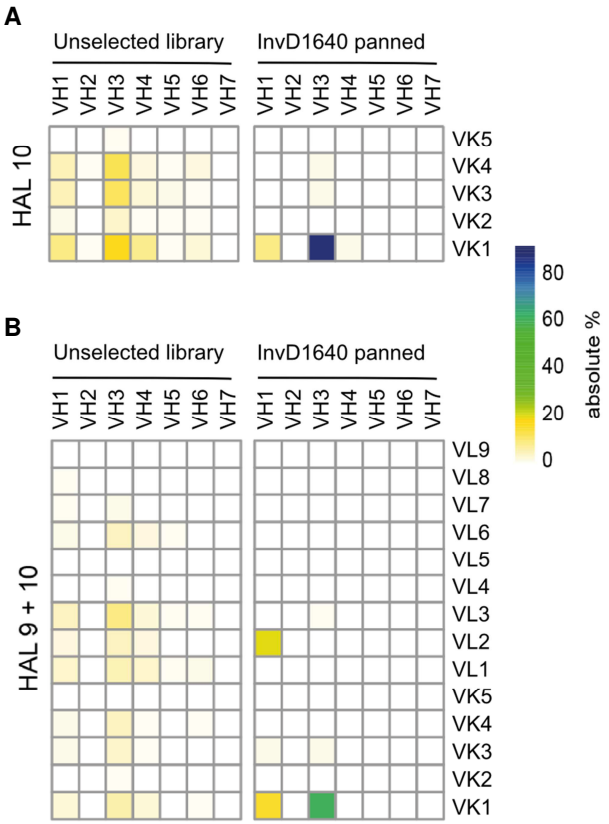

Figure 40: InvD targets VH3/VK1 subfamily antibodies. A) Heat map representation of the percentage abundance of (left) VH-VK gene pairs in HAL10 (kappa) library composed of 456 different clones and (right) VH-VK gene pairs obtained from the 3rd round of panning against InvD1640 using HAL10 library (based on sequence analysis of 57 clones). B) Heat map representation of the percentage abundance of (left) VH-VK/VL gene pairs in HAL9+10 (lambda + kappa) library composed of 1282 different clones and (right) VH-VK/VL gene pairs obtained from the 3rd round of panning against InvD1640 using HAL9+10 library (based on sequence analysis of 59 clones).

4.1.4.8 Pull down and MST experiments using recombinantly purified scFv-Fcs

To further confirm the interaction between InvD and VH3/VK1 subtype, pull down experiments using strep tagged InvD1640 with scFv-Fc fragments (different heavy/light chain subtype combinations) were performed. As a negative control ΔAD-InvD was used. Results confirm that InvD is specific towards the VH3/VK1 subfamily combination (Figure 41A). To determine the affinities between InvD and VH3/VK1 antibody, Cy5 labeled InvD was titrated with increasing concentrations of VH3/VK1-Fc or VH1/VL2-Fc antibodies. The obtained result confirmed the binding of InvD with VH3/VK1-Fc with an experimentally determined affinity constant (K_d) of 731 ± 261 nM (Figure 41B).

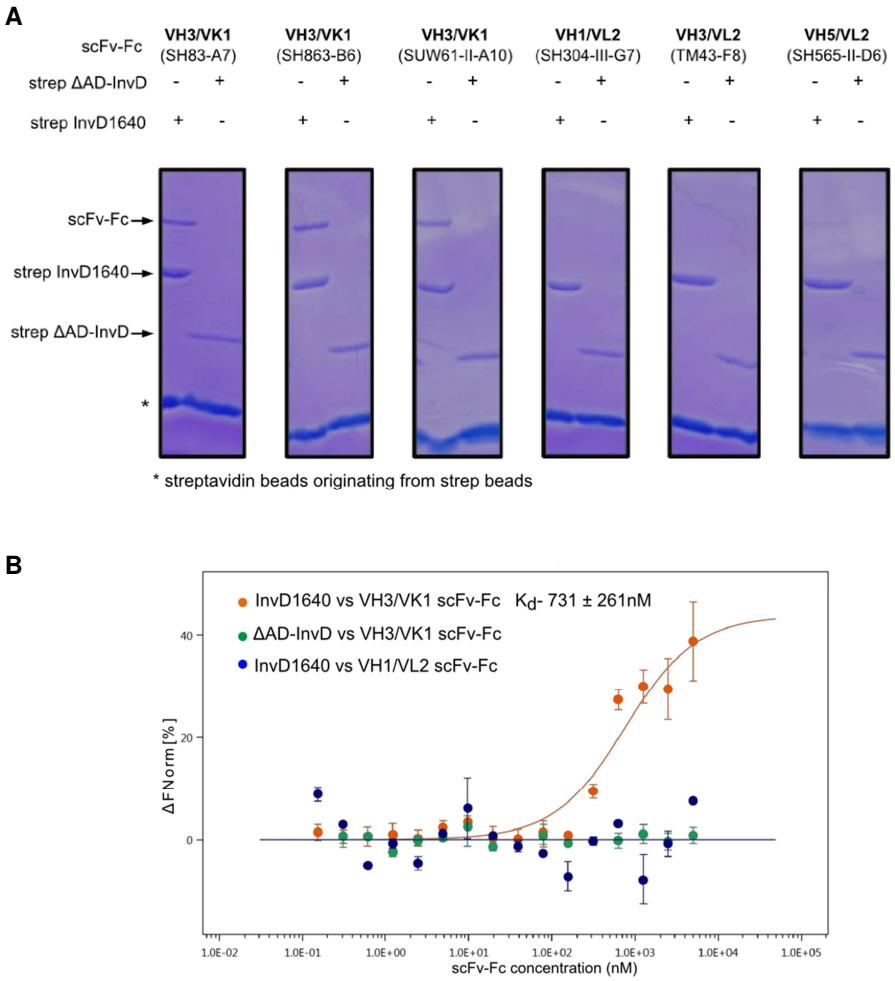

Figure 41: Interaction of InvD with scFv-Fc antibodies. A) Pulldown based interaction analysis of InvD and scFv-Fc antibodies with different heavy/light chain subtype combinations. B) Interaction of InvD1640/ΔAD-InvD with VH3/VK1 or VH1/VL2 subfamily scFv-Fc antibodies analyzed by MST. Data shown is the average of duplicate experiments.

4.2 InvE from IP31758 strain of *Y. pseudotuberculosis*

Strain IP31758 of *Y. pseudotuberculosis*is is responsible for Far East scarlet–like fever and also leads to toxic shock syndrome, which is not commonly seen for *pseudotuberculosis* infections (Eppinger et al., 2007). A novel Invasin, InvE has been recently identified in this strain. The C-terminal domain of InvE has a significantly different sequence from any other known Invasin. Thus, the aim of this part of work was to gain insights into the structure of the recently identified InvasinE.

4.2.1 Cloning, expression and purification of InvE

4.2.1.1 Bioinformatics analysis of InvE

Sequence analysis of InvE using a combination of BLAST-p, Pfam (A7FEC2/A0A0U1R0I0) and Signal-BLAST identified a putative signal peptide (SP in Figure 42) followed by a β-barrel domain comprising of residues 146-421, which is required for anchoring the protein to the outer membrane region of bacteria. The putative β-barrel domain is followed by 21 bacterial Ig-like (BIg) domains. BIg1-21 are capped by a predicted C-terminal domain (2619-2795 aa) with no similarity to known proteins. Based on these predictions, two constructs, InvE2547 (A2547-L2795) and InvE2448 (G2448-L2795) were designed (Figure 42).

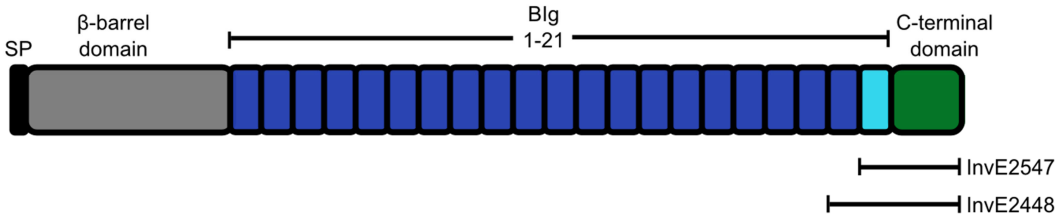

Figure 42: Domain architecture of *Y. pseudotuberculosis* InvE. Scheme of the domain organization of *Y. pseudotuberculosis* InvE with construct boundaries.

4.2.1.2 Cloning and expression

InvE constructs were amplified by PCR from the genomic DNA of *Y. pseudotuberculosis* IP31758 and cloned (3.2.1-3.2.7) into a modified pCOLA-Duet vector containing an N-terminal 6xHis tag (vector pVP009) and TEV cleavage site**.** After verification of inserts by sequencing (3.2.8), bacteria were transformed with the plasmid and tested for expression (3.3.1). Out of the two constructs, InvE2448 resulted in soluble protein whereas no expression was seen for InvE2547 (Figure 43).

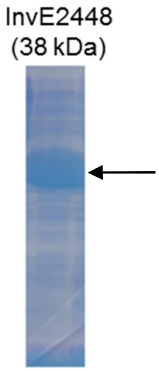

Figure 43: **Test expression of 6xHis_InvE2448.** Coomassie stained 15% SDS gel of test expression of 6xHis_InvE2448 construct. High yields of soluble protein were obtained by overnight incubation of cells at 18°C with 100 µM IPTG. Band corresponding to soluble InvE is marked with the black arrow.

4.2.1.3 Culturing and cell disruption

As described in section 3.3.2-3.3.3, 4 L of TB + kanamycin + chloramphenicol media was inoculated with the overnight culture of InvE2448 expressing bacteria. At an OD_{600} (optical density at 600 nm) of 0.7-0.8, expression was induced with 100 µM IPTG and cultures were allowed to grow at 18°C, 130 rpm for 21 h. After centrifugation and resuspension in lysis buffer, cells were homogenized.

4.2.1.4 Purification of 6xHis_InvE2448 for crystallization

After clearing of the lysate by centrifugation, affinity chromatography was performed (3.3.4). Elution fractions were pooled and TEV protease was added to remove the 6xHis tag. To remove the uncleaved material, reverse Ni-affinity was employed. To further purify the protein to homogeneity, size exclusion chromatography was performed (3.3.7) in 20 mM HEPES pH 7.4, 200 mM NaCl and 5 mM DTT (Figure 44).

Purified protein was identified as *Y.pseudotuberculosis* IP31758 InvE by tryptic digest and mass spectrometry. InvE2448 was concentrated to 25 mg/ml and flash frozen in liquid nitrogen.

4.2.1.5 Protease stability assay

To assess the stability and flexibility of InvE2448 in solution, 5 mg/ml of the protein was treated with 12 different proteases.

Apart from treatment with Proteinase-K and subtilisin, that resulted in stable sub-domain fragments, InvE2448 showed high stability with only minor degradation (Figure 45).

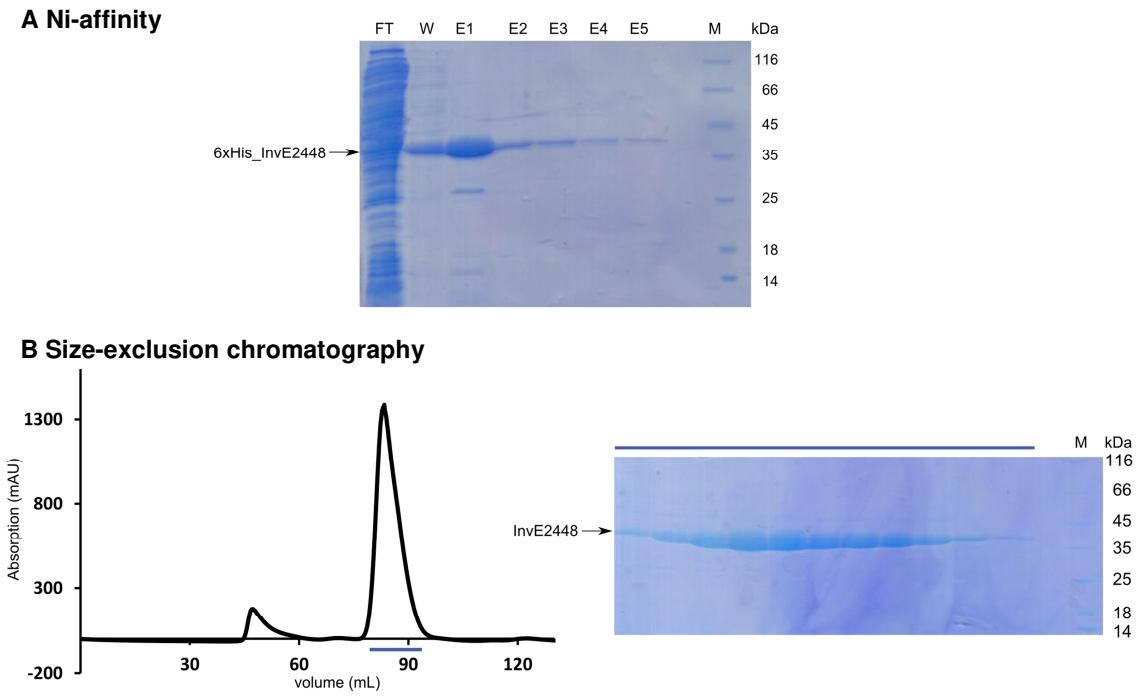

Figure 44: 6xHis_InvE2448 purification. A) Coomassie stained gel of InvE Ni-affinity chromatography. FT: flow through of the column; W: wash fraction (washing with 10 mM imidazole); E1-E5: elution fractions (with 100 mM imidazole); M: marker; B) Elution profile of the run on S200 16/60 column (left); Coomassie stained gel of eluted fractions. Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS PAGE. Absorbance of the protein is measured at 280 nm.

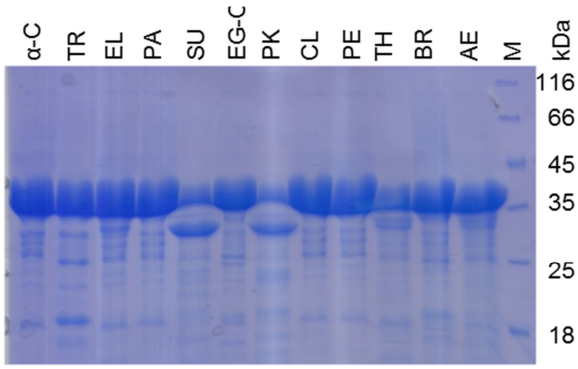

Figure 45: Stability of InvE2448 against proteases. InvE2448 (at a concentration of 5 mg/ml) was treated with 12 different proteases (α-C-α-Chymotrypsin, TR-Trypsin, EL-Elastase, PA-Papain, SU-Subtilisin, EG-C-Endoproteinase Glu-C, PK-Proteinase-K, CL-Clostripain, PE-Pepsin, TH-Thermolysin, BR-Bromelain, AE-Actinase, at a concentration of 0.01 mg/ml) and incubated for 1.5 h at 37°C.

4.2.2 Crystallization and structure solution of InvE

4.2.2.1 Crystallization of InvE

To gain insight into the structure of the adhesion domain of InvE, InvE2448 (without tag) was used for initial crystallization screening at 20°C and 4°C using commercially available screens. Initial crystals obtained with InvE2448 showed only weak diffraction. To improve crystal quality, a strategy of mutating surface entropic residues was applied. Using SERp server, four surface entropic residues (E2505A/Q2506A, K2646A, E2648A and

E2695A/K2696A) were identified. Out of the four mutants, K2646A was successfully expressed and purified in large scale as described for the wild type protein (Figure 46).

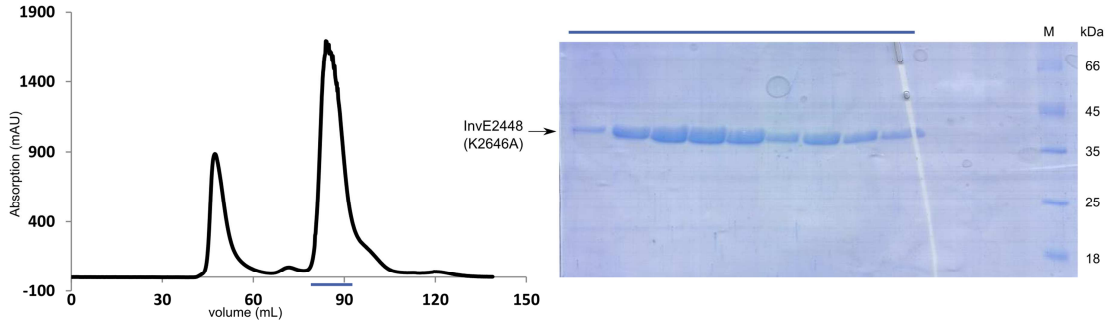

Figure 46: Purification of the surface entropy variant (K2646A) of InvE2448. Elution profile of InvE2448 (K2646A) on S200 16/60 column (left); Coomassie stained gel of eluted fractions (right). Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS PAGE. Absorbance of the protein is measured at 280 nm.

InvE2448 (K2646A) (without tag) was crystallized at a concentration of 18 mg/ml in the following condition- 0.12 M $MgCl_2$, 0.1 M TRIS pH 8.5, and 26% PEG8000. The crystals were grown in sitting drop at 4°C by mixture of 0.2 µL of concentrated protein with 0.2 µL of crystallization solution (Figure 47). For cryoprotection, the crystallization solution was supplemented with 10% PEG 400.

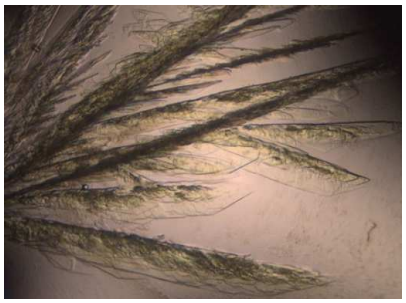

Figure 47: Crystals of InvE2448 (K2646A).

The native data set was collected to 1.7 Å at the SLS at X06DA/PXIII beamline. InvE C-terminal domain shows no significant similarity to other sequences in the protein database. Hence, initial molecular replacement trials were performed with Phaser MR by using individual Ig-like domains of InvA (PDB entry 1CWV (Hamburger, 1999) as a search template, but the approach was not successful.

Therefore, to solve the phase problem, a Se-Met labeled variant of InvE2448 (K2646A) was expressed and purified by Ni-affinity and size-exclusion chromatography (Figure 48). Expression and purification conditions were same as described for the native protein. Se-Met labeled InvE crystals grew under the same condition as the native crystals.

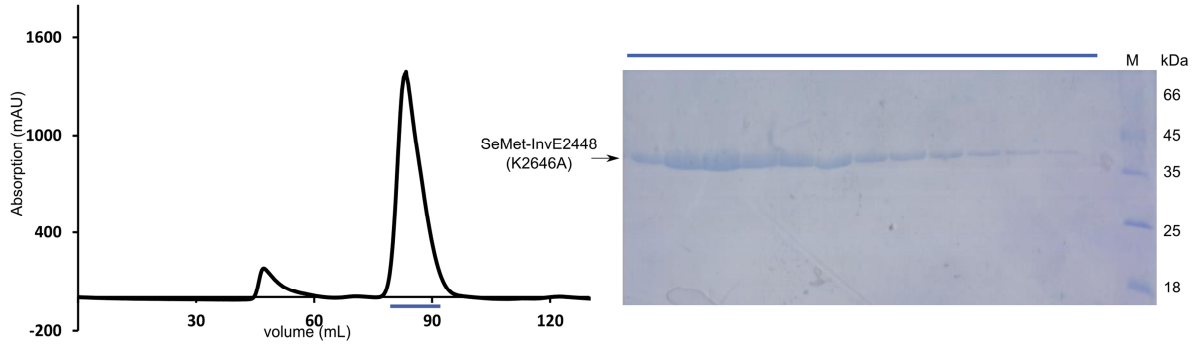

Figure 48: Purification of the Se-Met variant of InvE2448 (K2646A). Elution profile of InvE on S200 16/60 column (left); Coomassie stained gel of eluted fractions (right). Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS-PAGE. Absorbance of the protein is measured at 280 nm.

4.2.2.2 Data collection and structure solution

Data for Selenomethionine crystals were collected at SLS at X06DA/PXIII beamline to 2.3 Å resolution. The data was indexed, integrated and scaled using the XDS program package Se-Met InvE crystals belonged to space group $P2_1$. The structure was solved by Autosol program in Phenix. The model obtained after SAD phasing was subsequently used as the template to solve the native structure by molecular replacement. An initial model was automatically built using Phenix-autobuild and then manually rebuilt in WinCoot, followed by refinement in Phenix-refine. The final model was refined to 1.7 Å resolution. The overall electron density was of the high quality as shown in Figure 49. The statistics of data collection is shown in Table 26.

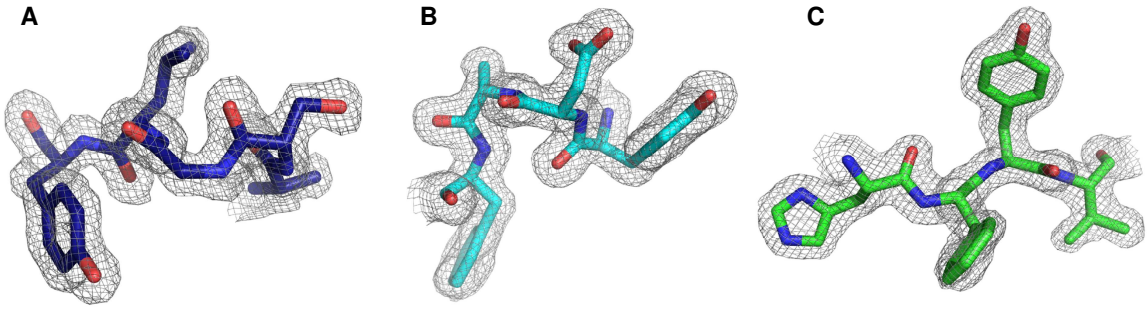

Figure 49: Quality of the electron density. $2F_o$-F_c electron density (contoured at 1 σ) in the A) Blg20, B) Blg21 and C) the C-terminal domain of InvE.

Table 26: Data collection and refinement statistics.

DATA COLLECTION	InvE	InvE (Se-Met derivative)
X-ray source	PXIII, SLS	PXIII, SLS
Wavelength [Å]	1.00000	0.97927
Resolution [Å]	46.4-1.7 (1.8-1.7)	46.3-2.3 (2.4-2.3)
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, b, c [Å]	70.26, 27.03, 88.95	70.13, 27.08, 88.83
α, β, γ [°]	90.00, 107.10, 90.00	90.00, 107.29, 90.00
R_{merge}**[%]**	7.7 (58.7)	10.9 (30.9)
R_{meas}**[%]**	8.4 (63.5)	11.1 (31.5)
CC1/2 [%]	99.9 (90.1)	99.9 (99.4)
‹I/σ(I)›	16.3 (3.2)	36.1 (15.1)
Completeness [%]	99.3 (99.7)	99.9 (100.1)
Unique reflections	35804 (5614)	27731 (3324)
Redundancy	6.5 (6.8)	27.8 (27.3)
REFINEMENT		
Resolution [Å]	46.4-1.7	
No. of reflections	35801	
R_{work}	0.161	
R_{free}	0.194	
No. of atoms	5722	
Avg. *B* factor protein [Å^2]	21.76	
R.m.s. deviations		
Bond length [Å]	0.003	
Bond angles [°]	0.557	
Ramachandran [%] (favored/allowed/ disallowed)	98.00/2.00/0.00	

4.2.3 Crystal structure of InvE

The overall structure of the construct is rod-like and displays a three domain architecture formed by two N-terminal domains Big20 (aa 2448-2546), Big21 (aa 2547-2660) and the adhesion domain (AD, aa 2661-2795) (Figure 50).

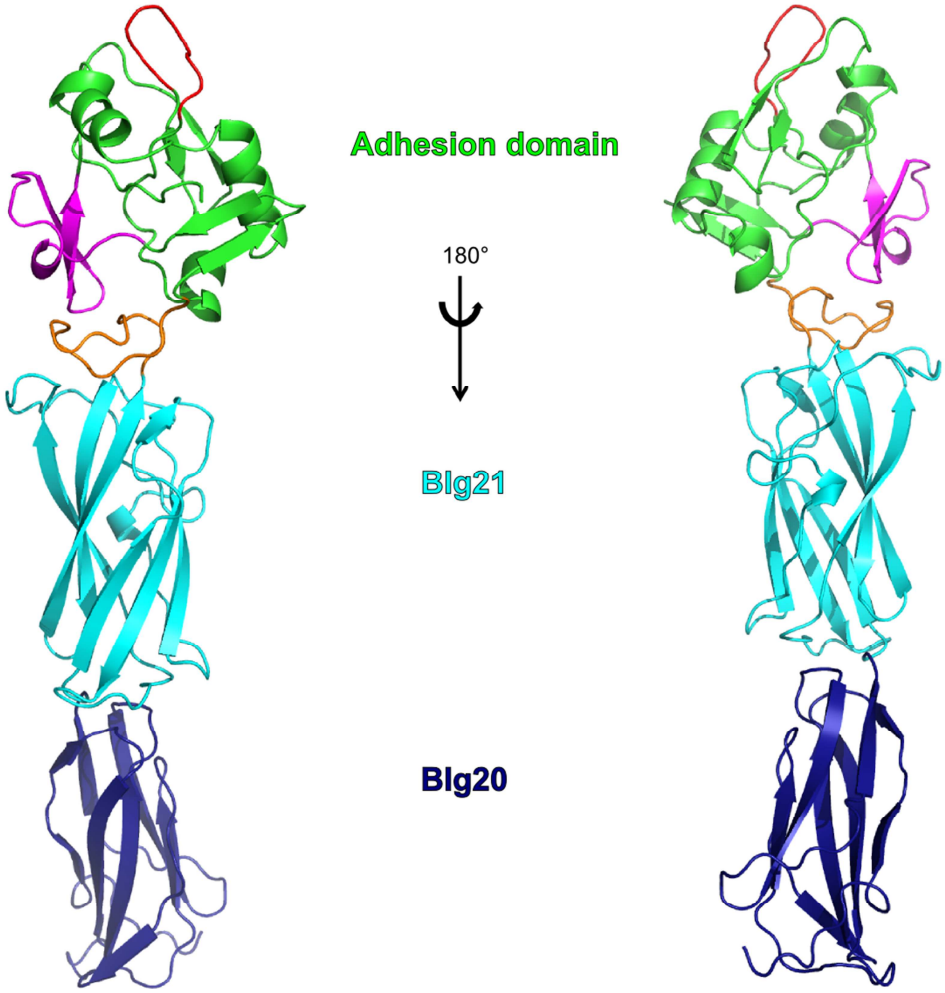

Figure 50: Crystal structure of *Y. pseudotuberculosis* IP31758 InvE. Cartoon model of the crystal structure of InvE2448. InvE2448 consists of the N-terminal Big20 (dark blue), Big21 (cyan) and the C-terminal adhesion domain (AD) (shown in green, magenta, orange and red; these elements are described in detail in 4.2.3.2) shown in two orientations.

4.2.3.1 SAXS structure of InvE2448

With the aim of understanding the structure of InvE in solution, SAXS data for InvE2448 was collected at BM29 of the ESRF in Grenoble, France. The calculated SAXS envelope is highly similar to the crystal structure of InvE2448 as shown in Figure 51A.

The chi value of 1.8 was obtained with CRYSOL on fitting the SAXS data with the crystal structure, which indicates that the model is in very good agreement with the calculated SAXS-envelope (Figure 51B).

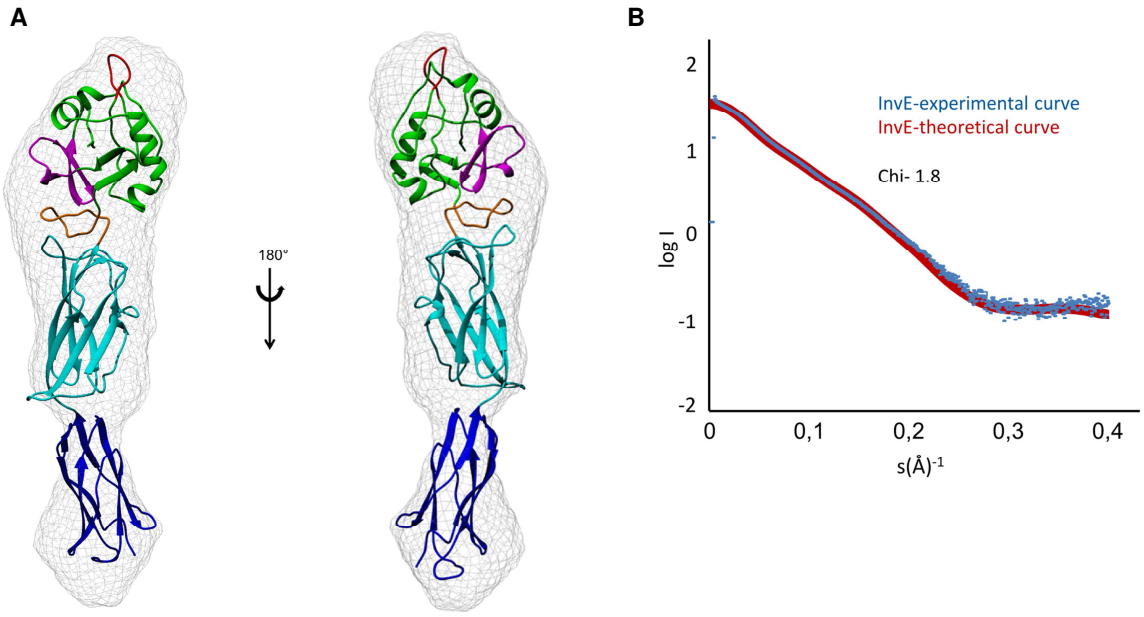

Figure 51: SAXS envelope of InvE2448. A) Rigid body fitting of the crystal structure into the *ab initio* determined SAXS envelope shown in two orientations. B) Fit of the rigid body model (red) with the experimental scattering (blue). Quality of the fit is expressed in terms of chi value.

4.2.3.2 Topology of InvE

BIg20/21 domains

Full length InvE consists of 21 BIg like domains, of these BIg1-20 share high sequence identity in the range of 56-98%, whereas BIg21 is significantly different from BIg1-20 (sequence identity 15-18%) as shown in Figure 52.

BIg20 of InvE adopts a two layer β-sandwich fold in which β-strands ABED form one sheet while the second β-sheet is formed by A'GFC (Figure 53). IgSF have previously been classified into V, C1, C2, I1 and I2-sets as described in section 4.1.3.3 and Figure 25. Presence of the A-A' strand with a kink formed by a cis–Proline (P2461), lack of C'' and presence of D strand in BIg20 are characteristics of the I1-set of IgSF (Figure 53). Even though it lacks strand C', it is classified in I1-set, since lack of C' has also been observed previously in other I1-set domains (Tan et al., 1998). Furthermore, two highly conserved cysteine residues forming a disulfide-bridge in the canonical Ig-fold are absent in BIg20 of InvE and instead replaced by V2474 (present on strand B) and L2526 (present on strand F).

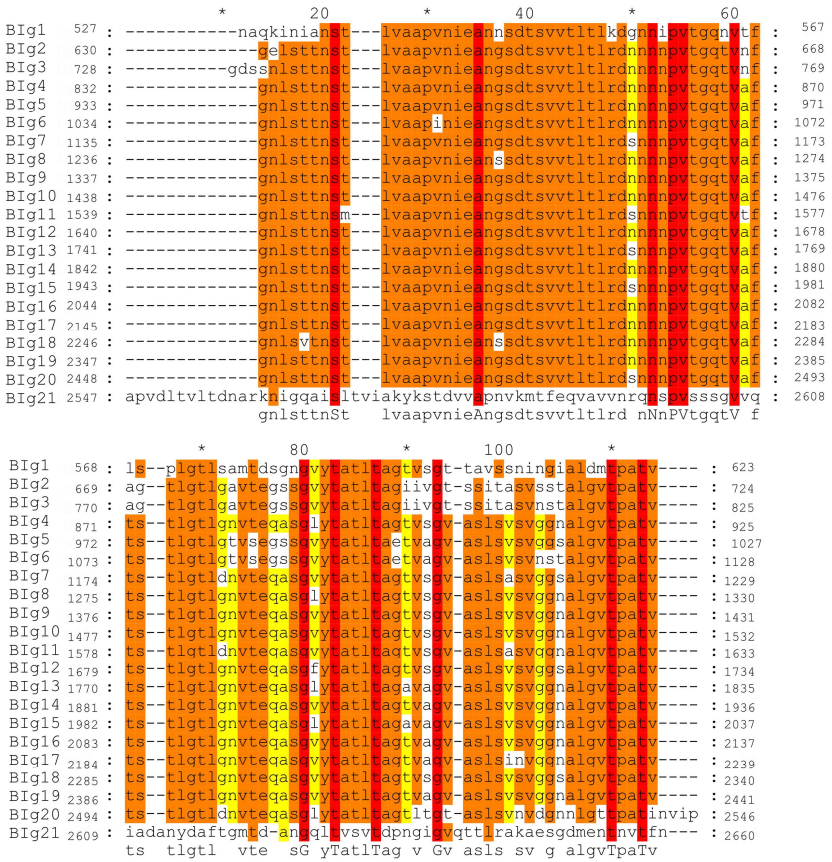

Figure 52: Sequence alignment of BIg1-21 of InvE. 100% conserved amino acids are marked in red, 80% in orange and 60% in yellow. Alignment has been generated with GeneDoc (Nicholas et al., 1996).

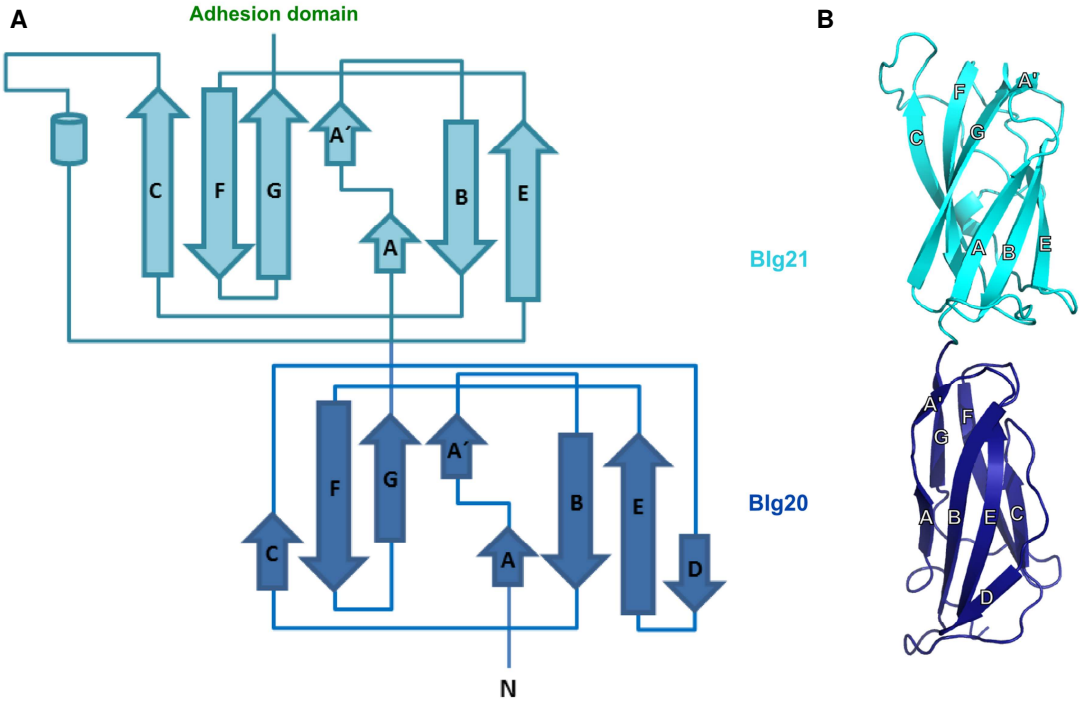

Figure 53: Overall structure of BIg20 and BIg21. A) Topology diagram of BIg20 (dark blue) and BIg21 (cyan) of InvE2448. (B) Cartoon representation of the BIg20/21 domains of InvE (β-strands are labeled).

Structural comparison of Big20 and Big21

As shown in Figure 54, the two domains can be superimposed with a rmsd of 2.6 Å. Although Big21 is topologically similar to Big20, there are significant differences between the structures of two domains.

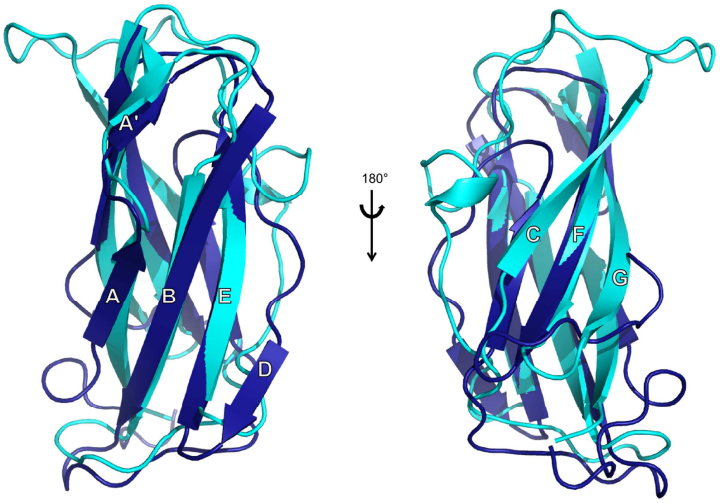

Figure 54: Structural overlay of Big20 and Big21. Superposition of Big20 (dark blue) and Big21 (cyan) shown in two orientations (β-strands are labeled).

Compared to Big20, Big21 has a longer C strand and a uniquely structured C-E interstrand loop along with a short helix that connects C strand with E (Figure 53). Conformation of the C-E interstrand loop is stabilized by three internal hydrogen bonds as shown in Figure 55.

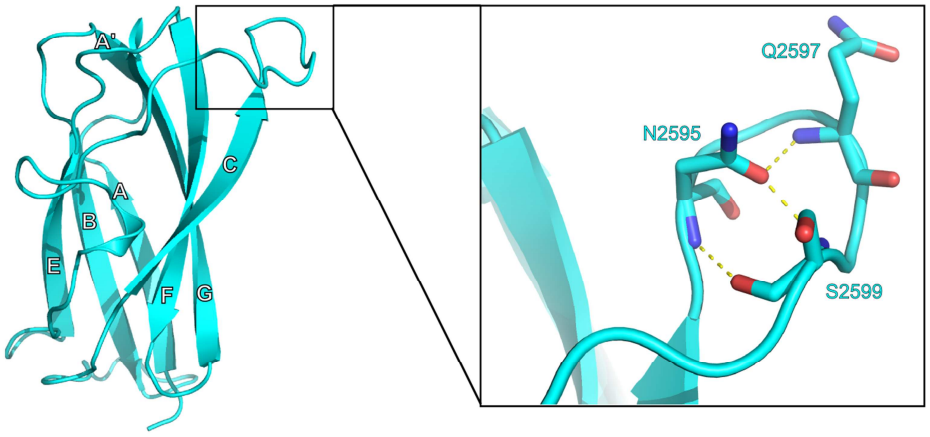

Figure 55: Stabilization of the C-E interstrand loop. Close up view of the C-E interstrand loop with residues stabilizing the loop are shown as sticks with oxygen atoms in red and nitrogen atoms in blue. Hydrogen bonds are shown as yellow dashed lines.

The conformation of this loop is unique to Big21 among structurally characterized Big's of InvE and InvA (Figure 56). In InvE, this loop is responsible for forming interactions between Big21 and the C-terminal AD as described in the context of the adhesion domain.

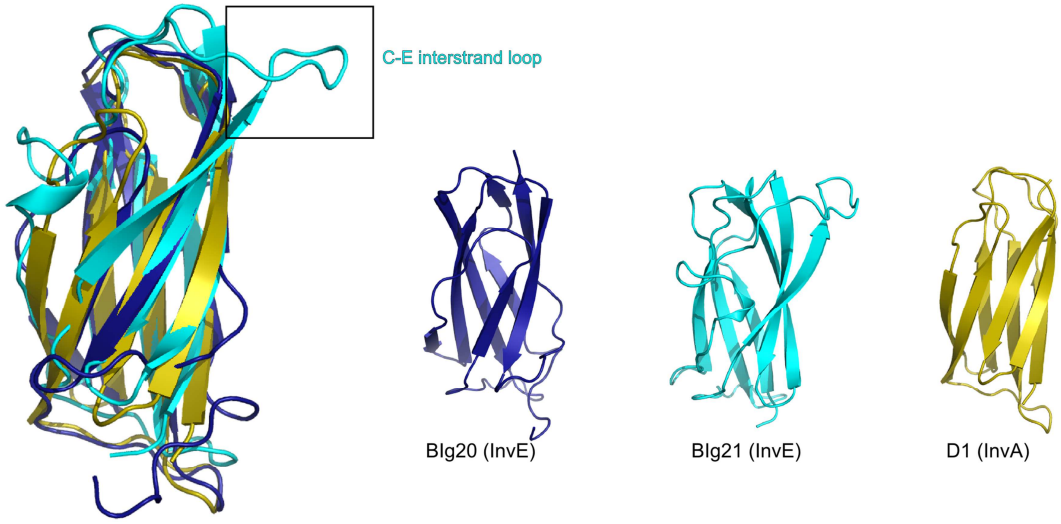

Figure 56: Conformation of the C-E interstrand loop. Structural overlay of BIg20 (dark blue), BIg21 (cyan) and D1 (InvA, in olive) (left). BIg domain structures of BIg20 (InvE), BIg21 (InvE) and D1 (InvA PDB ID-1CWV) (right).

BIg20/BIg21 interface

As calculated by PISA, the interface between BIg20/BIg21 spans an area of 319 Å^2 involving residues in the BIg20 A'-B and E-F interstrand loops and the BIg21 B-C interstrand loop (Table 27 and Figure 57A).

Table 27: PISA predicted BIg20-21 and BIg21-AD interacting interface.

	BIg20-BIg21 interface	**BIg21-AD interface**
N_{res}	17 (7 on BIg20 and 10 on BIg21)	29 (15 on BIg21 and 14 on AD)
N_{at}	63	121
Interface area (Å^2)	319.6	561.0
N_{HB}	2	6
N_{SB}	1	5
ΔG (kcal/mol)	-3.2	-8.2

N_{res}: indicates the number of interfacing residues; N_{at}: indicates the number of interfacing atoms; Interface area: half sum of the buried surface area; N_{HB}: number of hydrogen bonds in the interface N_{SB}: number of salt bridges in the interface; ΔG: gain in solvation energy.

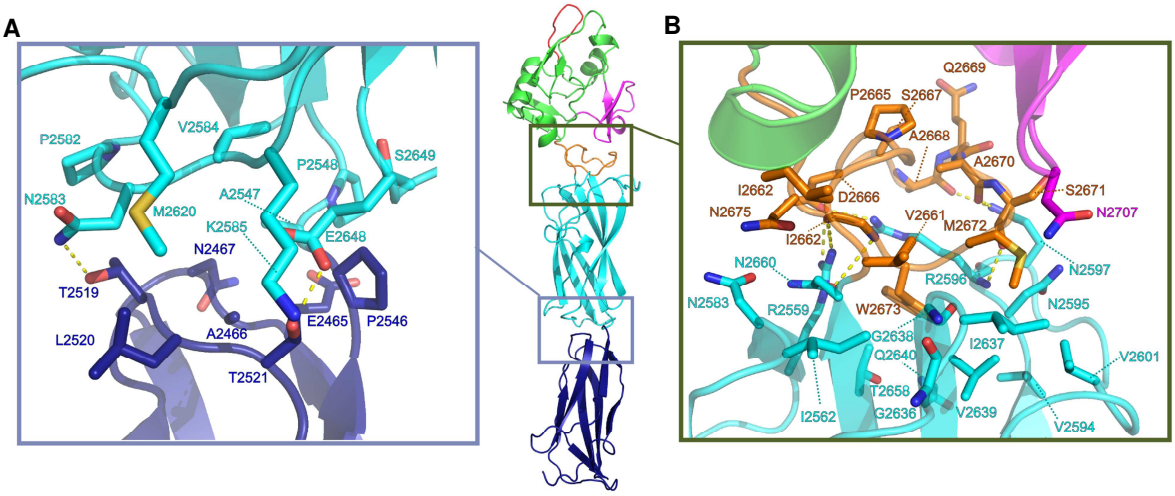

Figure 57: Blg20/21 and Blg21/AD interface. Detailed description of the inter-domain junctions of the A) Blg20–Blg21 and B) Blg21–AD. Residues responsible for forming interface contacts are labeled and shown as sticks with oxygen atoms in red and nitrogen atoms in blue. Hydrogen bonds are shown as yellow dashed lines.

Adhesion domain

The 1.7 Å crystal structure of InvE adhesion domain reveals resemblance to C-type lectin-like domains (CTLD) based on structure similarity search (discussed in detail in 5.2.2).

The canonical CTLD fold consists of two α-helices, that are oriented orthogonal to each other and two antiparallel β-sheets formed by strands β1+β5 and β2+β3+β4. The CTLD fold contains two disulfide bridges formed between β5 and α1 as well as between β3 and β5 (Figure 58).

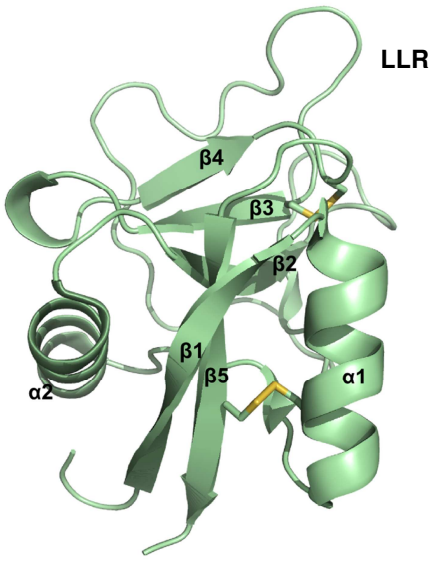

Figure 58: Cartoon representation of rat mannose binding protein (PDB ID-1MSB). This protein adopts a canonical C-type lectin-like domain fold. Secondary structure elements and long loop region (LLR) are labeled (Weis et al., 1991).

The long loop region (LLR) which is located between strands β2 and β3, plays an important role in receptor binding and also involved in Ca^{2+} binding in true C-type lectins. The CTLDs can be structurally subdivided into the canonical CTLD, which contains the LLR and the

hallmark motif „WIGL“, and into the compact CTLD, which lacks the LLR. Further division into sub-groups is based on the presence (long form) or absence (short form) of an N-terminal extension (presence of β1') (Torgersen et al., 1998; Zelensky and Gready, 2003, 2005).

Hence, based on the structural similarity of InvE AD to CTLDs, it is divided into three sections (Figure 59)–

i. a lasso-shaped loop structure (V2661-A2677, orange), which is termed as the “adaptor ring” (AR), since it forms a ring-shaped platform that connects Big21 with the residual sections of the AD,
ii. the “wedge domain” (WD, P2690-M2712, magenta)
iii. and the C-type lectin-like domain (CTLD, E2678-R2689 and F2713-L2795, green).

InvE CTLD comprises the canonical strands β1, β3 and β4. Residues usually forming strands β2 and β5 are located at the canonical position, however do not adapt to the β-strand geometry and are thus shown as coil instead of being displayed as β-strands in Figure 59 (also labeled in brackets for clarity). InvE CTLD contains two conserved Cysteines (C2722 and C2793) that link α1 to β5 with a disulfide bond but lacks the second conserved disulfide bridge present in between β3 and β5.

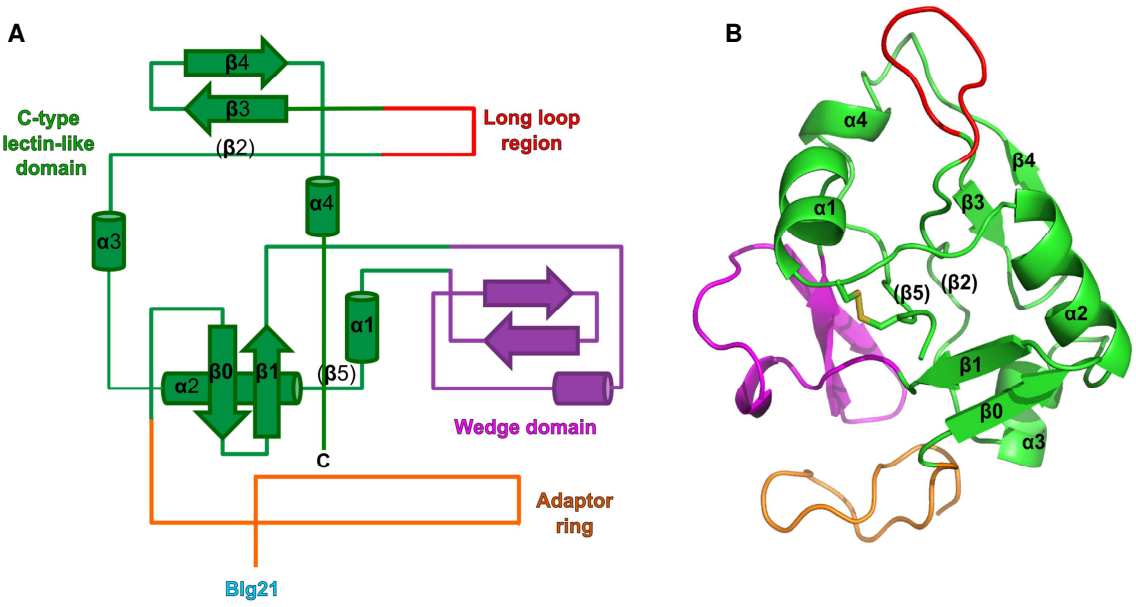

Figure 59: Overall structure of InvE adhesion domain. A) Topology of InvE AD. Different subdomains are labeled. B) Cartoon representation of InvE AD. Secondary structure elements are labeled. Disulfide bond is shown here as sticks with sulfur atoms in yellow.

In addition to two canonical α-helices α1 and α2, InvE CTLD comprises two additional α-helices, α3 and α4. Due to the presence of β0 as an additional N-terminal extension, it belongs to the long form of CTLDs. In addition, InvE lacks the hallmark „WIGL“ motif and its

LLR (red) only comprises less than eight residues and is thus classified as short. Thus, the CTLD of InvE belongs to the compact, long form CTLD sub-group.

AD and BIg21 domain interface

BIg21 and the AD span an extensive contact area interface of 561 $Å^2$. The interface is formed by a total of 29 residues from both the domains that make several contacts including five salt bridges. A tryptophan residue present in the AD (W2673) clamped between the C-E interstrand loop of BIg21, appears to play a major role in stabilizing the interaction between the two domains (Table 27 and Figure 57A).

4.3 YadA$_{pstb}$

YadA of *Y. enterocolitica* (ent) and *Y. pseudotuberculosis* (pstb) exhibit differences in their specificity of ECM substrate binding. This difference is due to a unique N-terminal 31 aa sequence, which is present in YadA$_{pstb}$ and absent in YadA$_{ent}$. The aim of this project was to provide the structural basis for the functional transition by solving the structure of YadA$_{pstb}$ and to compare it with YadA$_{ent}$ structure, which has already been solved previously (Nummelin et al., 2004).

4.3.1 Cloning, expression and purification of YadA$_{pstb}$

4.3.1.1 Bioinformatics analysis of YadA$_{pstb}$

In *Y. pseudotuberculosis*, YadA is expressed as 435 aa protein. Sequence analysis of YadA$_{pstb}$ (Uniprot ID-P10858) and its comparison with homologous YadA$_{ent}$ sequence (Uniprot ID-P31489) revealed that YadA$_{pstb}$ consists of the signal peptide (aa 1-25) that is cleaved off during transport to the outer membrane. This N-terminal sequence stretch is followed by the surface exposed head (aa 26-240) and stalk (aa 241-364) domains and a C-terminal membrane anchor domain (aa 365-435) (Heise and Dersch, 2006). To gain insight into structural differences between the head domain of YadA$_{pstb}$ and YadA$_{ent}$, YadA$_{pstb}$ construct comprising aa 26-253 was designed.

4.3.1.2 Cloning and expression of YadA$_{pstb}$

YadA$_{pstb}$ construct (aa 26-253) was amplified by PCR using pET28 plasmid containing YadA$_{pstb}$ (kindly provided by Prof. Petra Dersch) as a template and cloned (3.2.1-3.2.7) into a modified pCOLA-Duet vector containing an N-terminal 6xHis tag and TEV cleavage site (vector pVP009). After verification of the sequence of the insert (3.2.8), bacteria were transformed with the plasmid. YadA$_{pstb}$ resulted in soluble protein when tested for expression in *E. coli* (Figure 60).

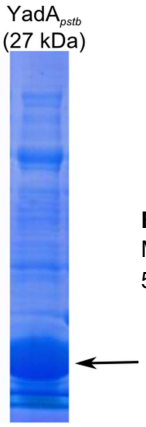

Figure 60: Coomassie stained 12% SDS gel of test expression of 6xHis YadA$_{pstb}$. Maximum yield of soluble protein was obtained by 3 h incubation of cells at 25°C with 500 µM IPTG. Band corresponding to soluble YadA$_{pstb}$ is marked with the black arrow.

4.3.1.3 Culturing and cell disruption

As described in section 3.3.2-3.3.3, 3 L of TB + kanamycin + chloramphenicol media was inoculated with the overnight culture of YadA$_{pstb}$. At an OD_{600} of 0.7, expression was induced with 500 µM IPTG and the culture was allowed to grow at 25°C, 130 rpm for 3 h. After centrifugation and resuspension in lysis buffer, cells were homogenized.

4.3.1.4 Purification of 6xHis tagged YadA$_{pstb}$

After clearing of the lysate by centrifugation, Ni-affinity chromatography was performed (3.3.4) (Figure 61A). YadA$_{pstb}$ containing fractions were pooled and incubated with TEV protease to remove the tag (3.3.5).

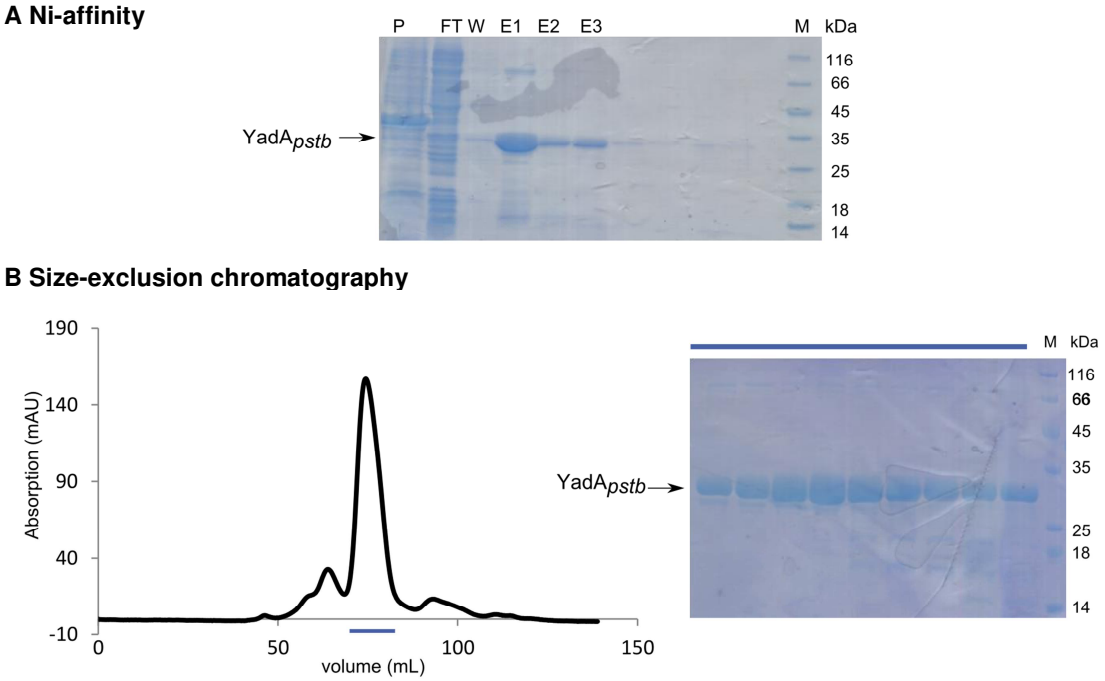

Figure 61: 6xHis YadA$_{pstb}$ large scale expression and purification. A) Coomassie stained gel of YadA$_{pstb}$ Ni-affinity chromatography. P: pellet; FT: flow through of the column; W: wash fraction (wash with 5 mM imidazole) E1-E3: elution fractions (with 100 mM imidazole); B) Elution profile of cleaved YadA$_{pstb}$ on S200 16/60 column (left); Coomassie stained gel of eluted fractions (right). Gel filtration retention volume indicates that YadA$_{pstb}$ eluted with an apparent molecular weight of ~75 kDa, suggesting elution as a trimer as expected. Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS-PAGE. Absorbance of the protein is measured at 280 nm.

Cleavage of tag was monitored by running uncleaved and cleaved samples on SDS-PAGE. After complete cleavage, reverse Ni-affinity was performed. Protein was further purified by size exclusion chromatography in 10 mM TRIS pH 8 and 300 mM NaCl (Figure 61B).

4.3.2 Crystallization and structure solution of YadA$_{pstb}$

4.3.2.1 Crystallization of YadA$_{pstb}$

To gain insight into the structure of YadA$_{pstb}$, purified YadA$_{pstb}$ (without tag) was used for initial crystallization screening at 20°C and 4°C using commercially available screens. 12 mg/ml of YadA$_{pstb}$ crystallized in 50% PEG 200, 0.1 M CHES pH 9.5 at 20°C (Figure 62). As the PEG200 concentration in the crystal condition was already sufficient for cryoprotection, the crystal was directly harvested from the drop, and flash frozen in liquid nitrogen.

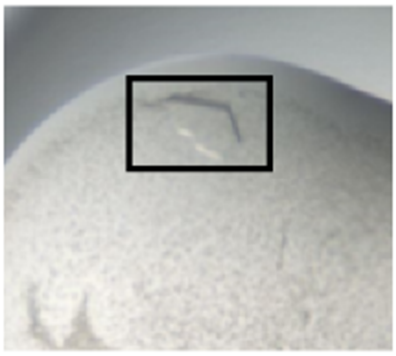

Figure 62: YadA$_{pstb}$ crystal. Crystal is marked with a black box.

4.3.2.2 Structure solution of YadA$_{pstb}$

The native data set (0.91841 Å) was collected at BESSY beamline 14.2. YadA$_{pstb}$ crystal diffracted to 2.3 Å. The data was processed using XDS and the structure was solved by Molecular Replacement using YadA$_{ent}$ (PDB ID-1P9H) (Nummelin et al., 2004) as a search model. Phenix Autobuild was used to build the model. Coot was used for further manual model building. Phenix refine was used for refining the structure. The statistics for data collection and refinement are shown in Table 28.

Table 28: Data collection and refinement statistics.

DATA COLLECTION	**YadA$_{pstb}$**
X-ray source	BESSY14.2
Wavelength [Å]	0.91841
Resolution [Å]	18.7-2.3 (2.4-2.3)
Space group	C2
Cell dimensions **a, b, c [Å]** **α, β, γ [°]**	 118.1,68.1,91.5 90, 94.119, 90
R_{merge}**[%]**	9.5 (36.0)
R_{meas}**[%]**	11.7 (45.7)
CC1/2 [%]	99.2 (87.5)
‹I/σ(I)›	8.2 (2.7)
Completeness [%]	92.0 (82.5)
Unique reflections	29835 (3168)
Redundancy	2.6 (2.2)
REFINEMENT	
Resolution [Å]	2.3
No. of reflections	29817
R_{work}	24.7
R_{free}	29.0
No. of atoms	3944
Avg. *B* factor [Å^2]	26.45
R.m.s. deviations	
Bond length [Å]	0.003
Bond angles [°]	0.774
Ramachandran [%] **(favored/allowed/** **disallowed)**	93.6/4.7/1.6

4.3.2.3 Crystal structure of YadA$_{pstb}$

YadA$_{pstb}$ crystallized as a trimer in space group C2. Like YadA$_{ent}$ (PDB ID-1P9H) (Nummelin et al., 2004b), YadA$_{pstb}$ consists of the head region that forms a left-handed parallel β-roll and the C-terminal domain comprising a short helix (Figure 63A). 12 N-terminal residues 26-37 are not visible in the crystal structure. However, the 31 aa insertion (53-83 aa) that is unique to YadA$_{pstb}$ is partially visible (12-16 aa out of 31 aa) (Figure 63B, C). In chain A, 53-64 aa, chain B, 53-69 aa, and in chain C, 53-67 aa are visible in the crystal structure.

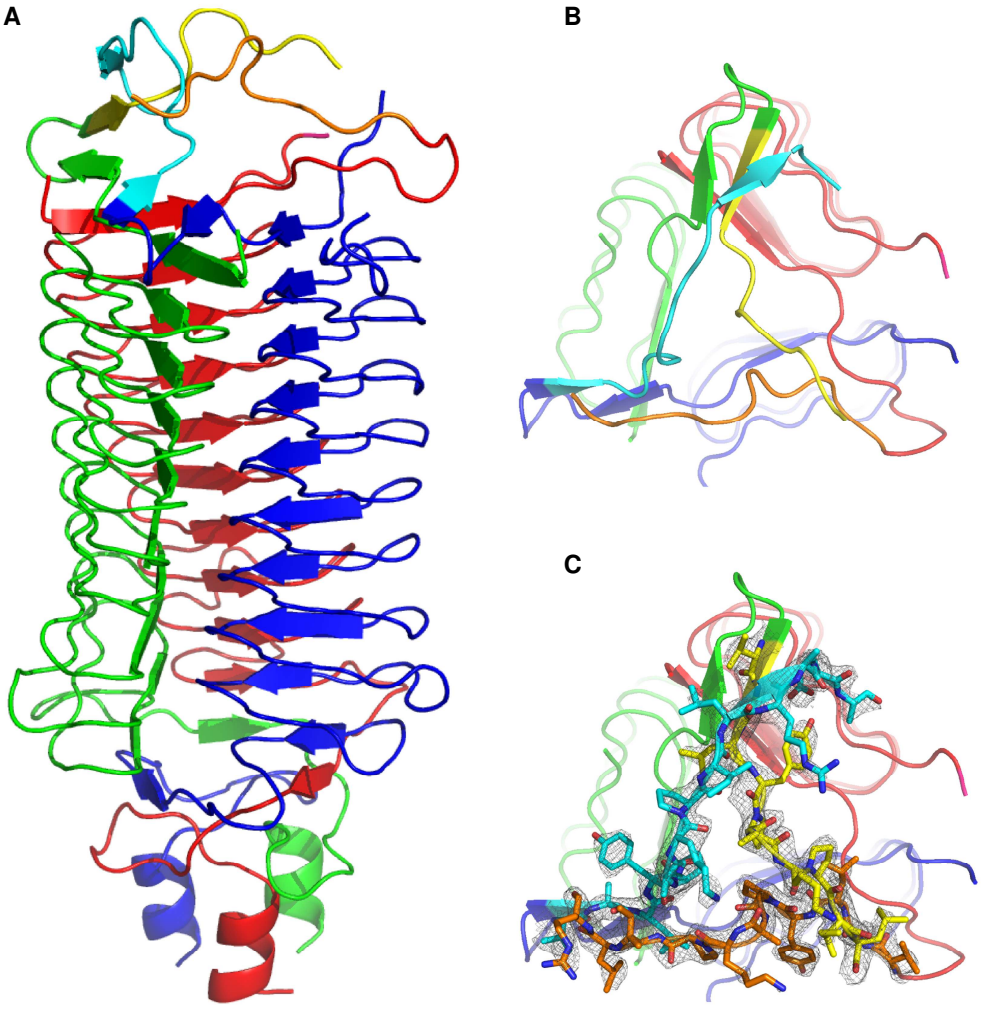

Figure 63: Crystal structure of *Y. pseudotuberculosis* YadA (YadA$_{pstb}$). A) Cartoon model of the trimeric crystal structure. Region that is unique to YadA$_{pstb}$ is colored in orange for the chain in red, cyan for the chain in blue, and yellow for the chain in green. B) Top view onto the tip with the structured parts of the 31 aa insertion, C) $2F_o$-F_c electron density (contoured at 1 σ) shown for the insertion region.

4.4 RovC

The aim of this part of work was to structurally and biochemically characterize the DNA binding protein RovC (Regulator of virulence associated with CsrC) from *Y. pseudotuberculosis* YPIII.

4.4.1 Cloning, expression and purification of RovC

4.4.1.1 Bioinformatics analysis of RovC

Full length RovC is expressed as 247 aa protein in YPIII strain of *Y. pseudotuberculosis*. BLASTp analysis of RovC sequence (Uniprot ID- A0A0H3B5N9) against the non-redundant protein database revealed that all of the hits were annotated as hypothetical proteins or proteins of unknown function (DUF2285). JPred (Cuff et al., 1998) and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) softwares were used to estimate the domain boundaries of RovC (Figure 64).

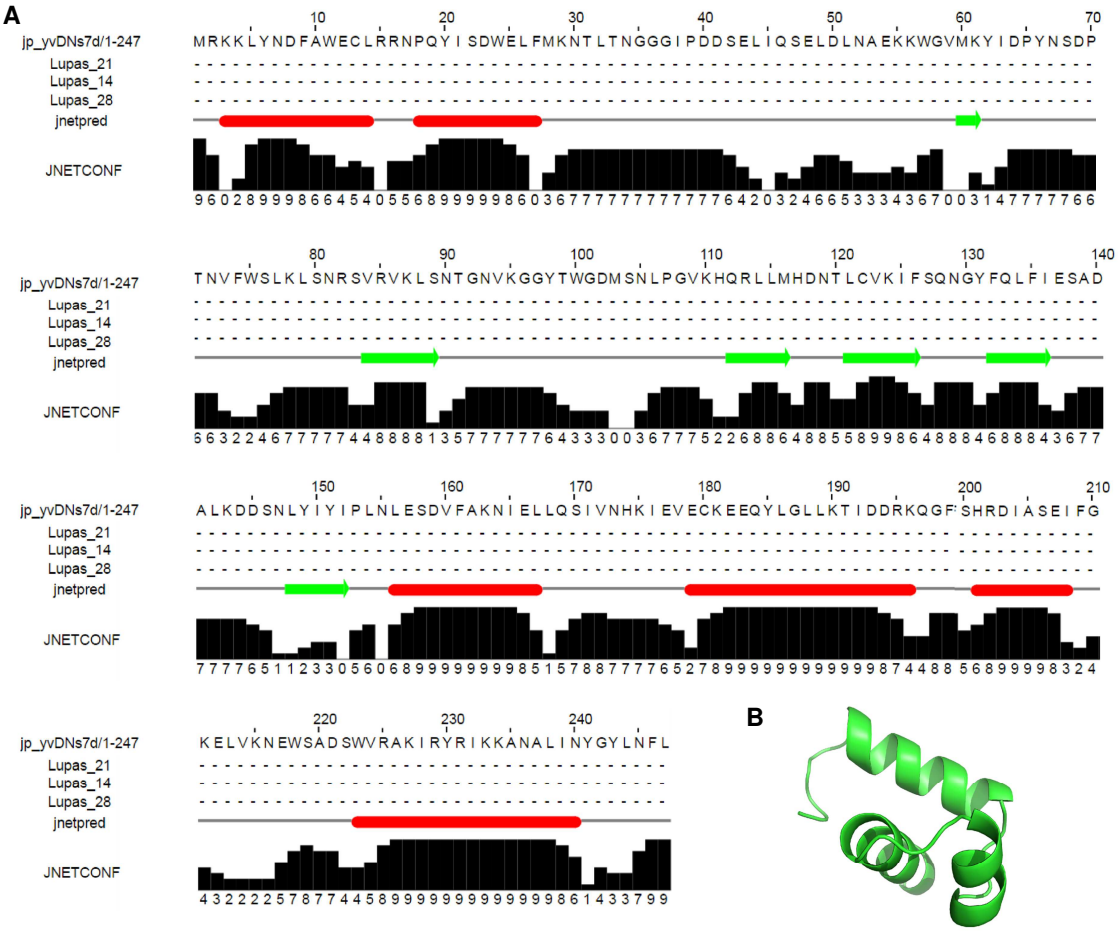

Figure 64: Bioinformatics analysis of RovC sequence. A) JPred secondary structure annotation of full length RovC (Lupas_21, Lupas_14, Lupas_28 are coiled coil predictions; JNetpred-secondary structure prediction-helices are shown in red and β-strands in green; JNETCONF-confidence estimated for the prediction). B) 3-D structure of the C-terminal region (185-247 aa) of RovC as predicted by the Phyre2 software.

The N-terminal region of RovC (1-154) has a predicted mixed α/β topology while the C-terminal region (155-247) is composed of α-helices (Figure 64A). Based on the tertiary structure prediction, a DNA binding domain was mapped in the C-terminal region (185-247 aa) that folds into a helix-turn-helix motif with 50% confidence (template PDB code-3KOR:D) (Figure 64B).

As the DNA binding domain was of special interest, three different N-terminal deletion constructs containing the predicted helix-turn-helix motif (D40-L247, S80-L247 and M103-L247) were designed.

4.4.1.2 Cloning and expression of RovC in *E. coli*

RovC constructs were amplified by PCR using genomic DNA of *Y. pseudotuberculosis* YPIII as a template and cloned (3.2.1-3.2.7) into a modified pCOLA-Duet vector with an N-terminal 6xHis tag and TEV cleavage site. After verification of inserts by sequencing (3.2.8), bacteria were transformed with the plasmid and tested for expression, but none of these constructs resulted in soluble protein. In another attempt, RovC (R2-L247; pVK14) fused with 6xHis_SUMO tag (kindly provided by Prof. Petra Dersch) was tested for expression in *E. coli*. Only a faint band for soluble RovC was observed, while most was found in the insoluble fraction.

Further attempts were made to increase the yield of soluble RovC. One strategy to increase the soluble expression is to lower the incubation temperature, which might lead to proper folding of the protein (Rosano and Ceccarelli, 2014). Moreover, ionic strength of the lysis buffer and presence of detergents also affect the solubility of proteins. Therefore; the following variables were investigated, either alone or in combination:

i) 16°C-20°C post-induction incubation temperature,

ii) 150 mM-500 mM NaCl in the lysis buffer,

iii) 0.1% Triton X-100 in the lysis buffer,

iv) dYT medium for growing of cells were tested.

RovC solubility was significantly improved on lowering the post-induction incubation temperature to 16°C and increasing NaCl concentration to 500 mM. Changing of LB medium to dYT medium also helped in improving the yield of soluble protein (Figure 65).

4.4.1.3 Culturing and cell disruption

As described in section 3.3.2-3.3.3, 4 L of dYT + kanamycin + chloramphenicol media was inoculated with the overnight culture of pVK14. At an OD_{600} of 0.6, expression was induced

with 120 µM IPTG and the culture was allowed to grow at 16°C, 130 rpm for 21 h. After centrifugation and resuspension in lysis buffer, cells were homogenized.

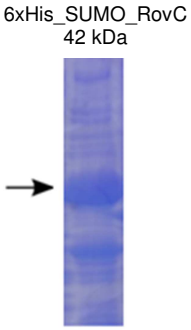

Figure 65: Coomassie stained 15% SDS gel of 6xHis_SUMO tagged RovC test expression. Maximum soluble protein yield was obtained at 16°C post induction incubation temperature with 120 µM IPTG in dYT medium. In the lysis buffer NaCl concentration was increased to 500 mM and 0.1% Triton X-100 was also added. Band corresponding to 6xHis_SUMO_RovC is marked with the black arrow.

4.4.1.4 Purification of RovC

After clearing of the lysate by centrifugation, affinity chromatography was performed. Due to the improved solubility at higher ionic strength, NaCl concentration was increased to 500 mM in all purification buffers. Protein was further purified by size exclusion chromatography in 100 mM TRIS pH 8, 500 mM NaCl, 5 mM DTT and 5% glycerol. A 31 kDa band copurified with the 42 kDa 6xHis_SUMO_RovC (Figure 66).

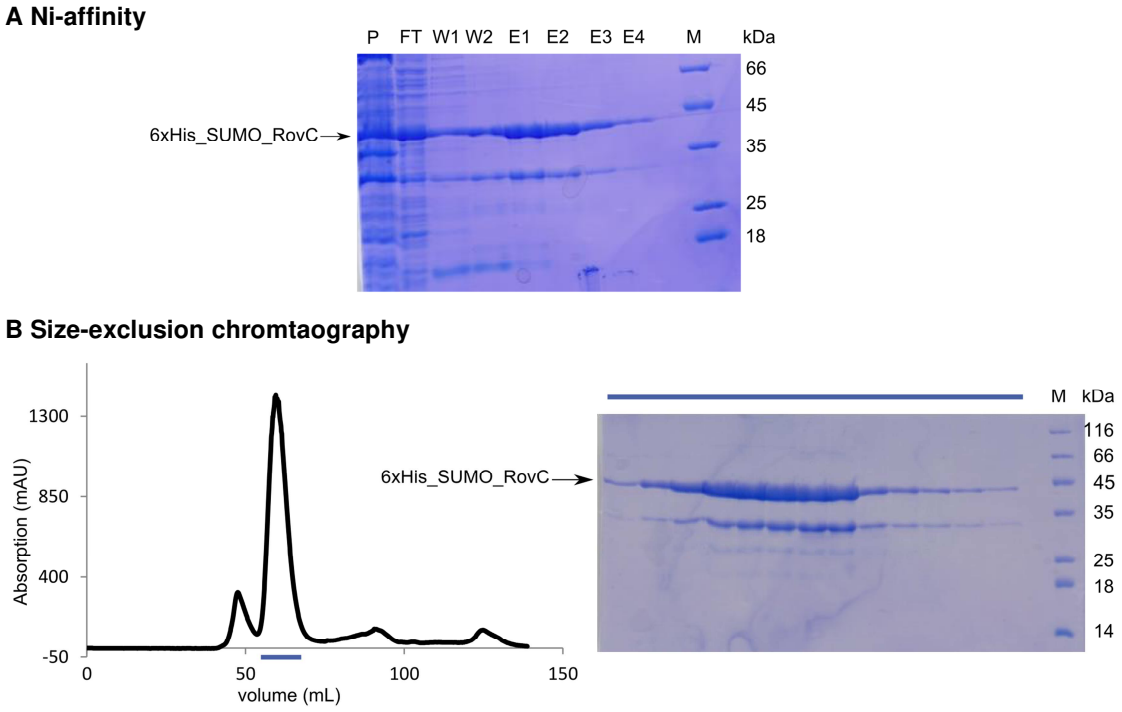

Figure 66: 6xHis_SUMO_RovC purification. A) Coomassie stained gel of RovC Ni-affinity chromatography. P: pellet; FT: flow through of the column; W1-W2: wash fraction (washing with 10 and 15 mM imidazole respectively); E1-E4: elution fractions (with 100 mM imidazole); M: marker; B) Elution profile of the run on S200 16/60 column (left); Coomassie stained gel of eluted fractions. Along with 42 kDa band (6xHis_SUMO_RovC), an additional band at 31 kDa was also observed (right). Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS PAGE. Absorbance of the protein is measured at 280 nm.

Mass-spectrometry analysis confirmed that both bands (42 and 31 kDa) belonged to RovC. To confirm the N-terminal sequence of the smaller fragment, the band was analyzed by Edman degradation that identified the sequence as I-E-A-H-R-E-Q-I-G, corresponding to the C-terminal end of SUMO tag. The results from Edman degradation reveal a partial degradation of the tag, resulting in two species, the 6xHis_SUMO_RovC and the degraded variant, partially lacking the 6xHis_SUMO tag. Due to this partial degradation, a homogeneous sample was not obtained.

To circumvent this problem, RovC (M1-L247) was cloned into a modified pCOLA-Duet vector (pPS41) with N-terminal 6xHis_SUMO tag and TEV cleavage site. RovC was expressed under the same condition as described above.

Protein was first purified by Ni-affinity chromatography. Elution fractions were pooled and TEV protease was added to remove the 6xHis_SUMO tag. To remove the uncleaved material, reverse Ni-affinity was performed. Cleavage with the highly specific TEV protease also resulted in cleavage of the partially degraded species observed before, resulting in a homogenous protein. Final step of purification was performed by size exclusion chromatography under the conditions described above (Figure 67). The overall yield of pure RovC was about 5 mg/L culture.

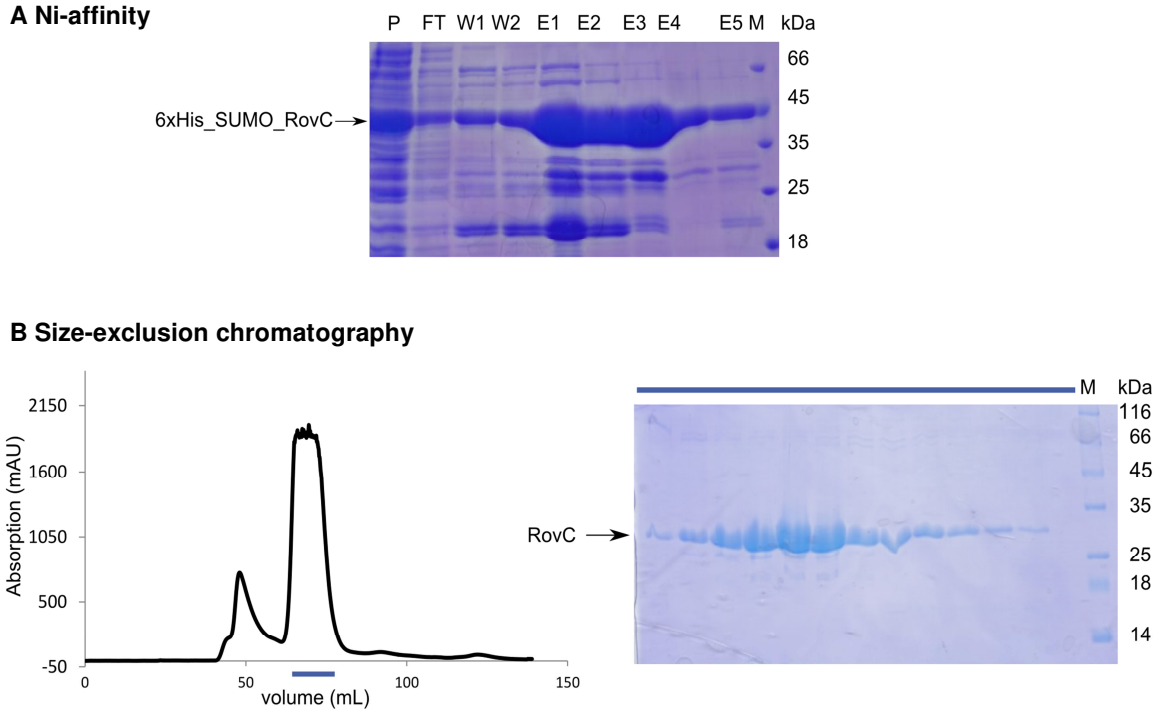

Figure 67: 6xHis_SUMO_TEV_RovC purification. A) Coomassie stained gel of RovC Ni-affinity chromatography. FT: flow through of the column; W1-W3: wash fraction (washing with 5, 10 and 15 mM imidazole respectively); E1-E5: elution fractions (with 100 mM imidazole); M: marker B) Elution fractions were pooled, incubated with TEV and loaded on size exclusion column. Elution profile of the run on S200 16/60 column (left); Coomassie stained gel of eluted fractions (right). Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS PAGE. Absorbance of the protein is measured at 280 nm.

4.4.2 Crystallization and structure solution

4.4.2.1 Crystallization of RovC

Crystallization screening with RovC (M1-L247) was performed at 20°C and 4°C using commercially available screens (anion, JCSG I-IV). RovC crystals were obtained under several different crystallization conditions. Most promising crystals were obtained with 0.1 M TRIS pH 8.5, 3.5 M sodium formate at 4°C with 4 mg/ml of RovC. For cryoprotection, reservoir solution was supplemented with an additional 15% glycerol. These crystals diffracted to 4 Å in-house (Figure 68A). Crystals were further optimized by designing a 96 well grid screen around that condition. Single crystals were obtained under the following condition-0.1 M TRIS pH 8.3, 3.7 M sodium formate with 3 mg/ml of protein at 4°C (Figure 68B).

After optimization, crystals diffracted to 3.8 Å in-house. Several of those crystals were dissolved in protein buffer (100 mM TRIS pH 8, 500 mM NaCl, 5 mM DTT, 5% glycerol) and loaded on SDS PAGE to confirm the identity of crystals. As a control, RovC (in solution) was also loaded on the gel. The protein band was sent for mass spectrometry analysis which confirmed that the band in the gel belongs to RovC (Figure 68C).

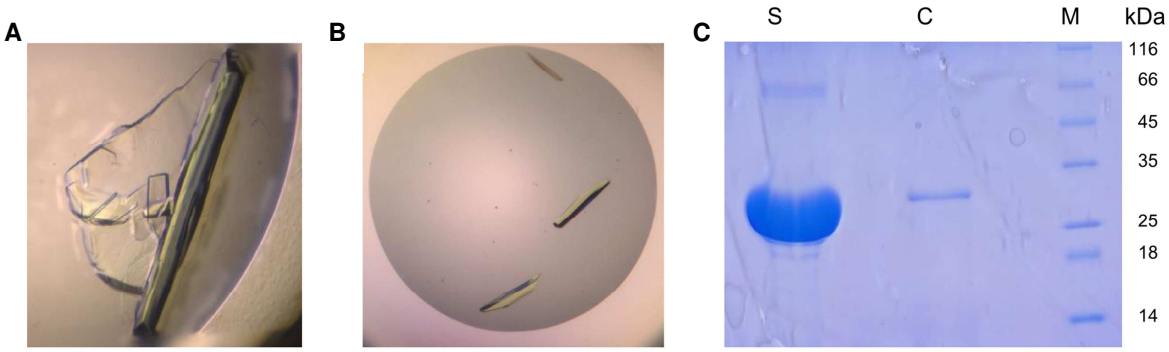

Figure 68: RovC crystals. A) Initial hit B) Optimized crystals C) SDS PAGE analysis of crystals (S=protein in solution, C=protein in crystals).

RovC shares no sequence homology with any other protein in the database, hence to solve the phase problem, a Se-Met labeled variant of RovC was expressed and purified by Ni-affinity and size exclusion chromatography (Figure 69). Expression and purification conditions were same as described for the native protein.

Se-Met labeled RovC crystals grew under the same condition as the native crystals.

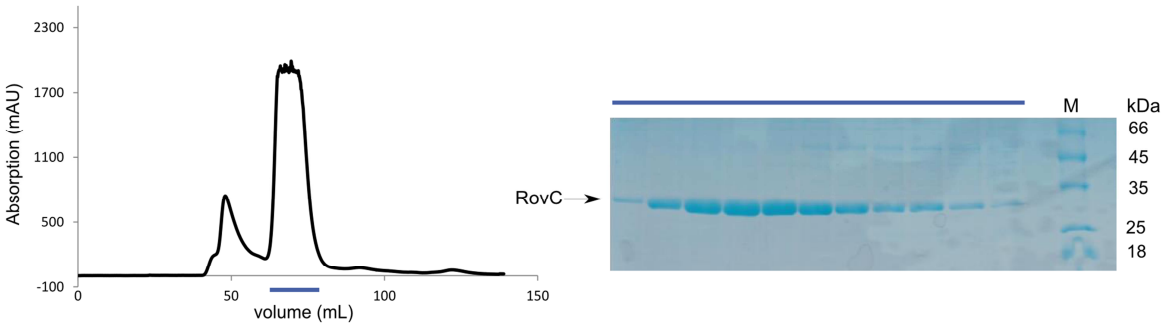

Figure 69: Purification of the Se-Met variant of RovC. Elution profile of Se-Met RovC on S200 16/60 column (left); Coomassie stained gel of eluted fractions (right). Blue bar represents the corresponding elution fractions in the chromatogram that were analysed by SDS PAGE. Absorbance of the protein is measured at 280 nm.

4.4.2.2 Data collection and structure solution

The data for Selenomethionine (peak wavelength-0.979287 Å) crystal was collected at beamline P11 at PETRA III at DESY. The data was indexed, integrated and scaled using the XDS program package. Se-Met RovC crystals diffracted to 3.0 Å resolution and belonged to space group *R*32:H. With one molecule in the asymmetric unit, the Matthews coefficient V_M equals 2.58 Å^3/Da, corresponding to a solvent content of 52%. The structure was solved by Autosol program in Phenix. Total of four Se sites (out of five) were located in one molecule with a figure of merit of 0.37. After an initial round of model building using the AutoBuild module in Phenix, a total of 132 out of 247 residues (63 residues in the N-terminal region and 69 residues in the C-terminal region) were built with R_{work}/R_{free} of 34%/41% followed by manual building in Coot. The quality of the electron density was however not sufficient to build the complete model.

With the aim to find crystals with better diffracting properties, further screening of RovC against other commercially available screens (Nucleix, Morphius and Midas) were performed. Crystals were obtained in the following condition- 0.2 M KCl, 0.01 M $MgSO_4$, 0.01 M MES pH 5.6, 10% PEG400 at 20°C with 6 mg/ml of RovC. For cryoprotection, reservoir solution was supplemented with an additional 20% glycerol. Data for native crystals were collected at beamline P11 at PETRA III at DESY. Crystals diffracted to 2.3 Å resolution and belonged to space group P321. With one molecule in the asymmetric unit, the Matthews coefficient V_M equals 2.21 Å^3/Da, corresponding to a solvent content of 44%. The structure was solved by Molecular Replacement using Phaser MR by using the N- and C-terminal domains of the previously build partial model of RovC separately as search models. Phenix AutoBuild was used to build the model. In this crystal form, additional electron density was visible, into which residues were built manually in Coot. Even though the optimized crystals improved the overall quality of the electron density, the quality of the additional electron density was not sufficient to assign a sequence, and has thus only been built as poly-Ala model. The statistics of the data collection is summarized in Table 29.

Table 29: Data collection and refinement statistics.

DATA COLLECTION	**RovC**	**RovC SeMet**
X-ray source	DESY PXIII	DESY PXIII
Wavelength [Å]	1.033200	0.979287
Resolution [Å]	33.44 - 2.30 (2.38 - 2.30)	19.40 - 3.00 (3.10 - 3.00)
Space group	P321	R32
Cell dimensions		
a, b, c [Å]	97.28, 97.28, 46.04	97.67,97.67,156.03
α, β, γ [°]	90, 90, 120	90, 90, 120
R_{merge}**[%]**	8.3 (56.6)	17.8 (267.5)
R_{meas}**[%]**	9.6 (65.7)	18.1 (271.9)
CC1/2 [%]	99.6 (69.4)	99.9 (85.8)
‹I/σ(I)›	9.7 (2.3)	21.46 (2.2)
Completeness [%]	99.5 (99.9)	99.7 (99.5)
Unique reflections	11309 (1095)	11112 (1039)
Redundancy	4.0 (4.0)	30.9 (31.0)
REFINEMENT		
Resolution [Å]	33.44 - 2.30	
No. of reflections	11306	
R_{work}	22.6	
R_{free}	28.1	
No. of atoms	1713	
Avg. *B* factor [Å^2]	39.16	
R.m.s. deviations		
Bond length [Å]	0.008	
Bond angles [°]	0.930	
Ramachandran [%] (favored/allowed/ disallowed)	96.1/3.9/0.0	

4.4.2.3 Crystal structure of RovC

RovC consists of two distinct folded domains (Figure 70), the N-terminal domain, comprising residues 1-135 (shown in green) and the C-terminal domain comprising residues 183-247 (shown in teal). The N-terminal domain consists of three α-helices that are packed against three anti-parallel β-strands. These secondary structure features are connected by flexible loops that are not well resolved in the crystal structure. The C-terminal domain that adopts a helix-turn-helix motif is well defined by its electron density. Residues 136-176 (linker, shown in light brown) were built as poly-alanine backbone, because the side-chains were not visible. The overall quality of the electron is shown in Figure 71.

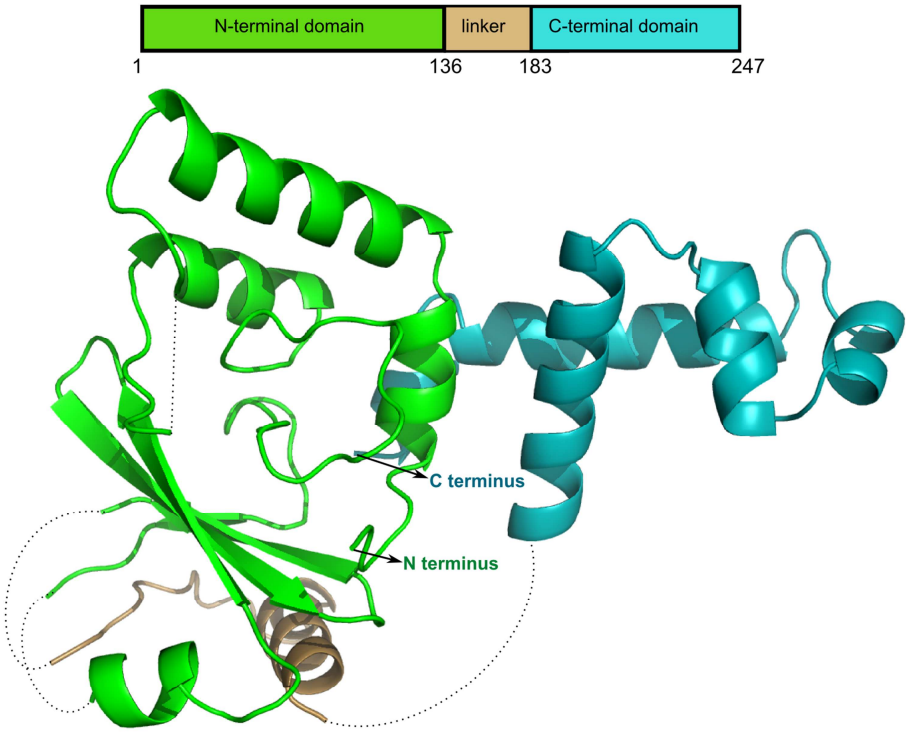

Figure 70: Crystal structure of RovC. Domain boundaries of RovC (top) and cartoon representation of RovC (bottom). RovC consists of the N-terminal domain (shown in green), linker (shown in light brown) and the C-terminal domain (shown in teal). Flexible regions that cannot be built are shown as black dotted lines. N and C terminus are labeled.

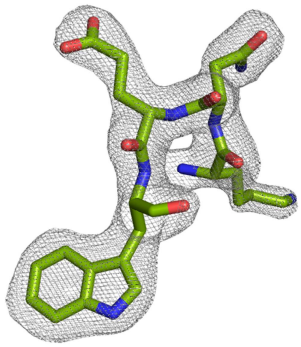

Figure 71: Quality of the electron density. A representative portion of the electron density map ($2F_o$-F_c contoured at 1 σ) showing the quality of the electron density.

4.4.3 Strategies to alter crystallization behavior of RovC

The crystal structure contained regions that were not well resolved; therefore, the approach was to alter the crystallization behavior of RovC by either modifying surface properties of the protein or removing disordered protein regions. With the aim of obtaining new crystal forms or better resolution, three different strategies were applied.

The first approach was to clone, express and purify a set of surface entropy mutants of RovC. Using SERp server, three surface entropic residues (54-56EKK, 181-183KEE, 233-234KK) were identified. Out of the three mutants, K181A/E182A/E183A and K233A/K234A were successfully cloned, expressed and purified in large scale (Figure 72). Crystallization assays were set up but no new crystal form was obtained.

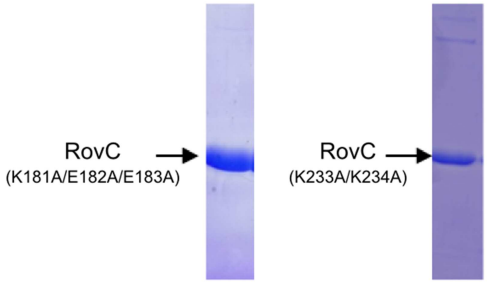

Figure 72: Purity of purified RovC SER mutants. Coommassie stained gel showing purity of mutants after size exclusion chromatography. These proteins were expressed and purified as described for the wild type RovC.

In-situ proteolysis was used as a second approach to rescue crystallization. Out of 12 different proteases, only EndoproteinaseGluC produced a stable subdomain fragment (marked by the dashed box in Figure 73). EndoproteinaseGlu-C was added to RovC at a concentration of 1:1000 and 1:10000; 1.5 h prior to setting up of crystallization screens. Although RovC crystallized in various conditions, it did not yield well diffracting crystals.

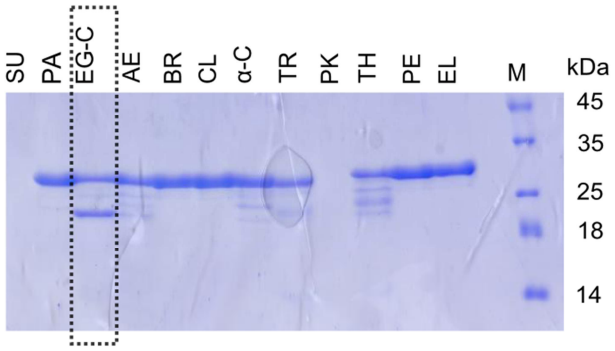

Figure 73: Treatment of RovC with different proteases. RovC at a concentration of 1 mg/ml was treated with 12 different proteases (SU-Subtilisin, PA-Papain, EG-C-EndoproteinaseGlu-C, AE-Actinase, BR-Bromelain, CL-Clostripain, α-C-α-Chymotrypsin, PK-Proteinase-K, TH-Thermolysin, TR-Trypsin, PE-Pepsin, EL-Elastase at a concentration of 0.01 mg/ml) and incubated for 1.5 h at room temperature.

As a third approach, reductive methylation of lysines was performed (Figure 74). Crystallization screening was performed with Lys-methylated RovC, however, this approach also did not yield good diffracting crystals.

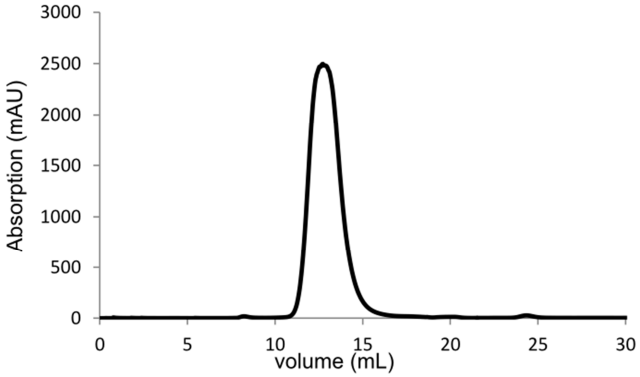

Figure 74: Lysine methylation of RovC. Elution profile of the run on S200 10/30 column. Absorbance of the protein is measured at 280 nm.

Even though different strategies were used, no improvement of the crystal quality compared to the crystals obtained with wild type potein could be obtained.

4.4.4 Oligomeric state of RovC

4.4.4.1 Multi Angle Light Scattering (MALS) of RovC

A fully cleaved RovC (without 6xHis_SUMO tag), which has a molecular weight of 29.2 kDa, eluted between 60 and 80 ml on S200 16/60 column (Figure 67), suggestive of an oligomeric state. An oligomeric state would also be in line with the previous observation, that partially truncated 6xHis_SUMO_RovC, lacking the N-terminal 6xHis-tag, still co-eluted with full-length 6xHis_SUMO_RovC from Ni-affinity resin.

In order to determine the exact oligomeric state of RovC, size exclusion chromatography followed by multi-angle laser light scattering (SEC-MALS) experiment was performed. Molecular mass calculation revealed that the molecular weight is 173 kDa, which is close to the theoretical mass of the hexamer (175.2 kDa) revealing that RovC assembles as a highly stable hexamer in solution (Figure 75), since no additional peaks were observed in the elution profile, apart from the hexamer peak.

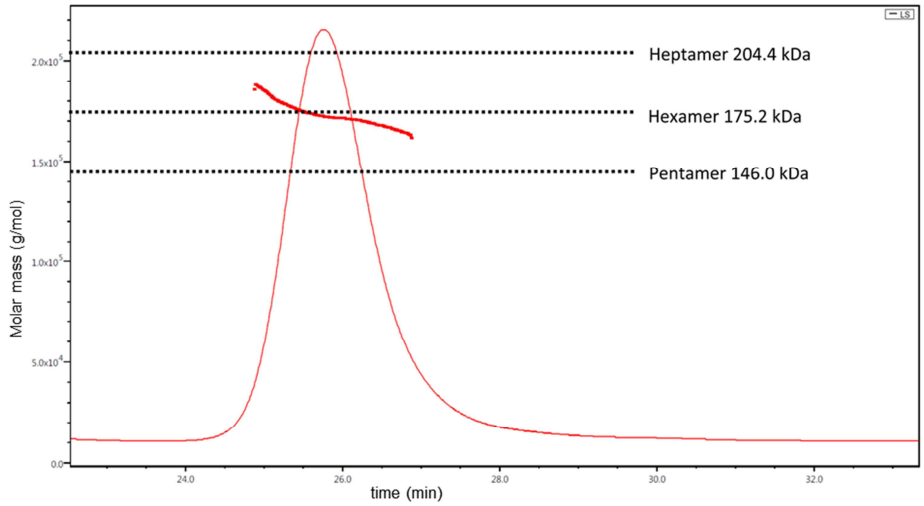

Figure 75: Light scattering chromatogram of RovC. SEC-MALS profile of purified RovC as observed from the MALS signal.

4.4.4.2 Hexamer in the crystal structure

In the presented crystal structure, the asymmetric unit contains only one RovC molecule per AU (Figure 76A); hence, first dimer (Figure 76B) and then potential hexameric assemblies were generated by crystallographic two and three fold (32) symmetry operators. Two independent hexameric structures, ring and triskelion shaped were observed in the crystal structure (Figure 76C and D).

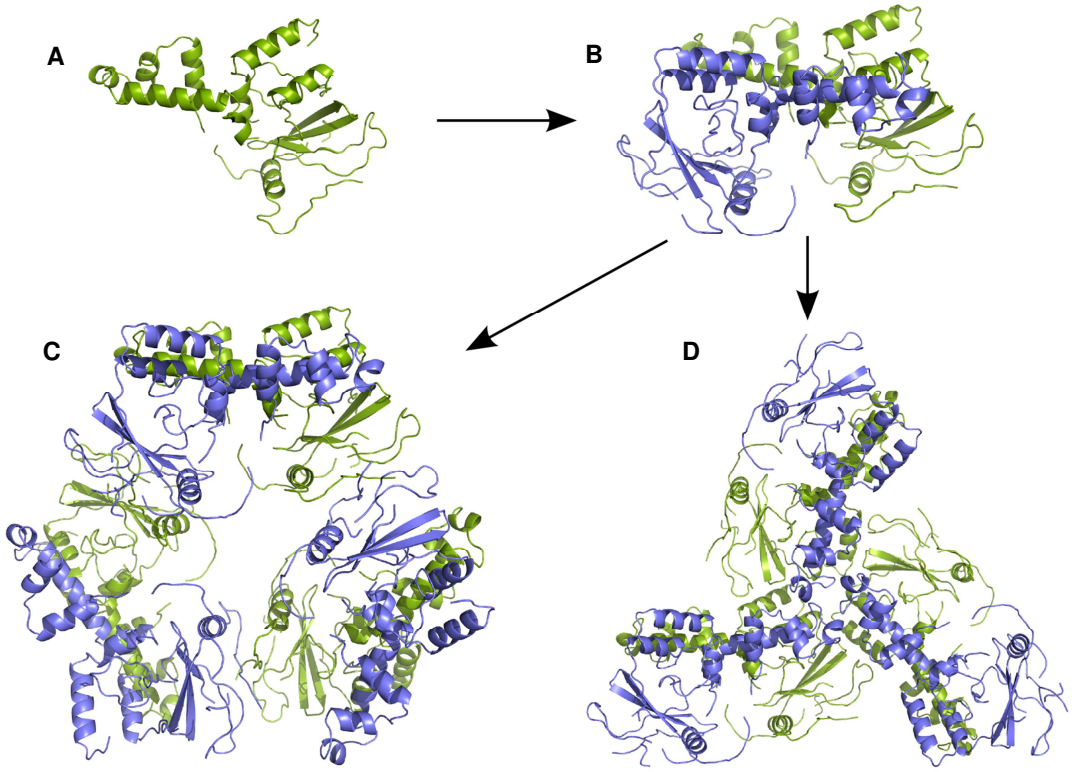

Figure 76: RovC assemblies generated using the symmetry mates. A) monomer present in the asymmetric unit. B) RovC dimer, C) ring shaped hexamer and D) triskelion shaped hexamer generated by applying crystallographic symmetry.

4.4.4.3 Small Angle X-Ray Scattering analysis of RovC

To determine how RovC oligomers assemble in solution, SAXS experiment was performed. SAXS envelope was constructed without symmetry (P1) and also with symmetry restriction (P32 as seen in the crystal structure). Ab initio calculated envelopes were very similar (Figure 77A, B). Fitting of the calculated scattering curve of RovC with the ring shaped hexamer resulted in a chi value of 1.27 which indicates a very good fit to the experimental scattering data (Figure 77C).

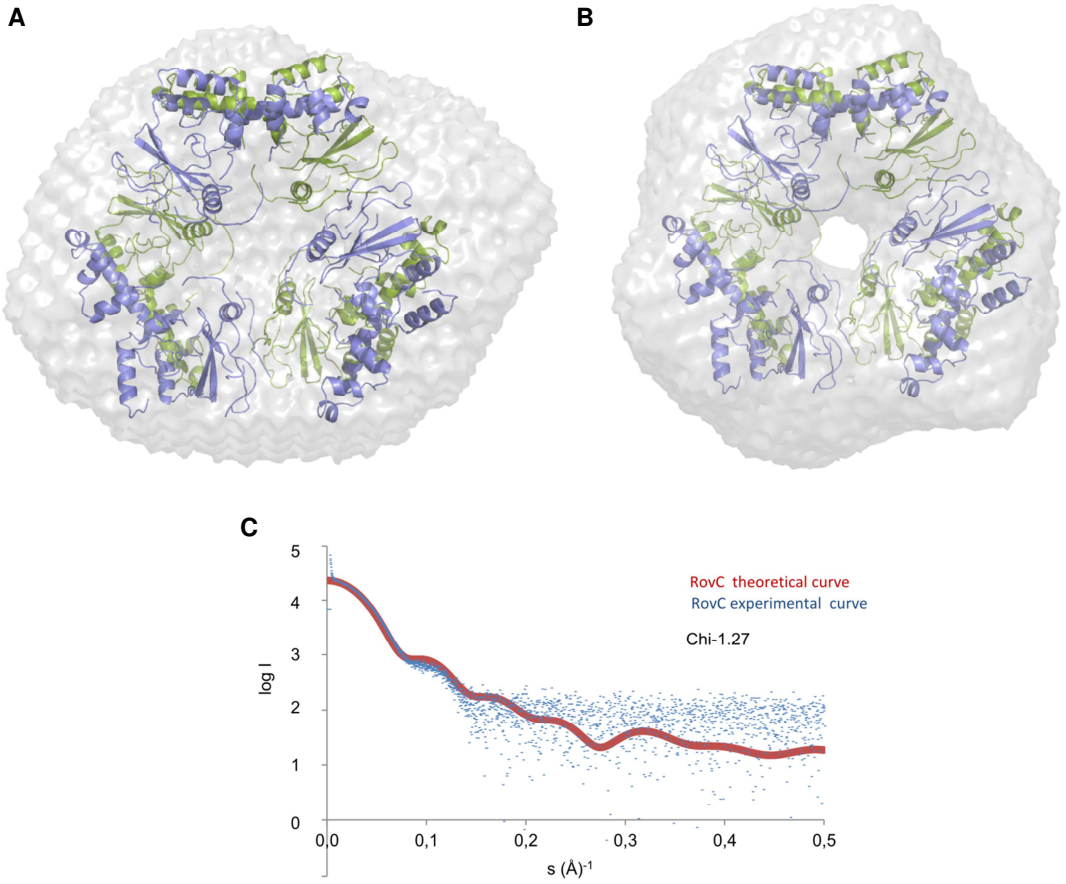

Figure 77: SAXS analysis of RovC. Rigid body fitting of the hexamer into the ab initio determined SAXS envelope calculated A) with no symmetry imposed B) with P32 symmetry imposed. C) Fit of the rigid body model (red) with the experimental scattering (blue). Quality of the fit is expressed in terms of chi value.

4.4.4.4 Structure-guided mutational analysis of RovC

SAXS data confirmed the ring shaped structure of RovC. To further validate the ring-shaped model, a series of point mutations were constructed based on the crystal structure that were divided into Set I, Set II and Set III categories (Figure 78).

1) Mutations that disrupt the dimeric interface (Set I) – The dimer interface is formed mostly by residues 231-246 of the C-terminal domain (Figure 78 and Figure 79). The A237E and G242E mutations that might disrupt the dimeric interface of RovC due to replacement of the small/hydrophobic amino acid with a negatively charged side chain were designed.

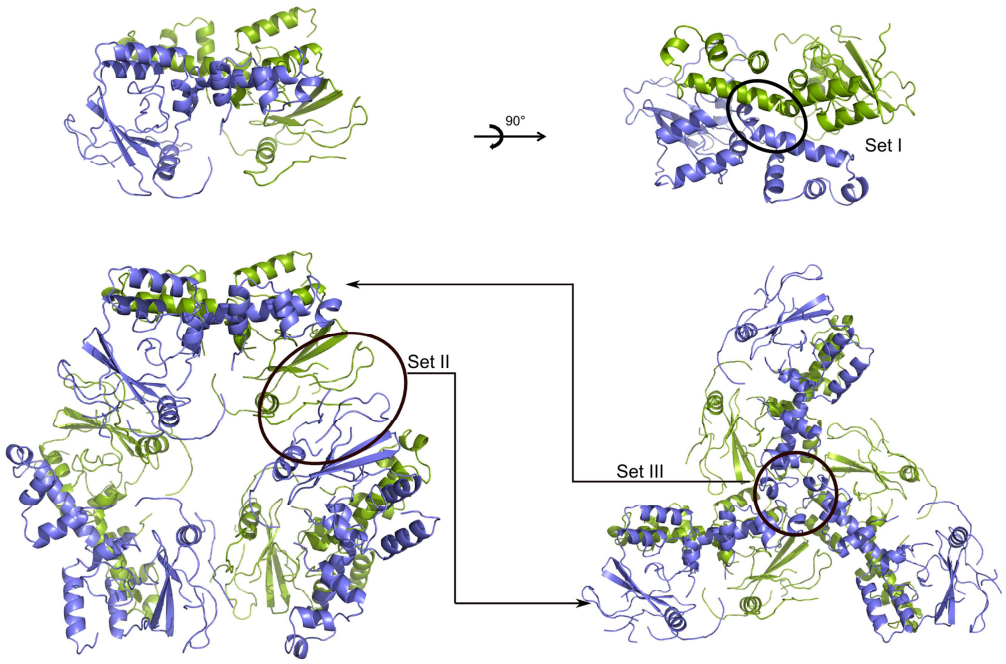

Figure 78: Scheme of different mutations introduced in RovC. Regions where mutations are introduced are marked with the black circle.

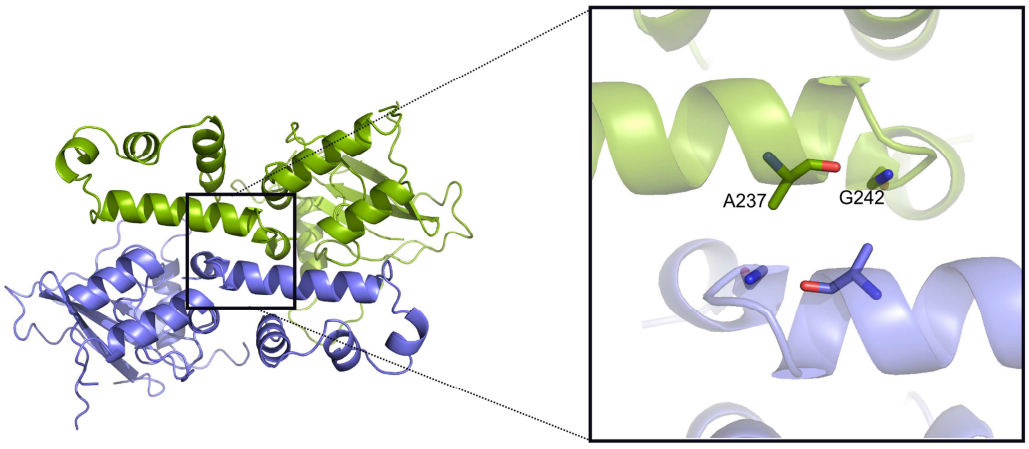

Figure 79: Close-up view of the dimer interface of RovC. Residues that were mutated (Set I) are labeled (only in 1 molecule) and shown as sticks with oxygen atoms in red and nitrogen atoms in blue.

2) Mutations that disrupt the ring shaped hexamer (Set II) - In the ring shaped hexamer; the interface formed between two dimers is contributed by residues 142-180 that is not well defined in the crystal structure. Based on the secondary structure prediction (Figure 64), residues 147-152 form a β-strand. Even though the side chains could not be assigned for the additional density in the crystal structure, one section adopts a β-strand conformation that is likely to correspond to region 147-152. Thus, the I150P, I150P/Y151P mutations that might disrupt the interface in the ring shaped hexamer by destabilizing β-strand structure were designed. These mutations however are solvent

exposed in triskelion shaped structure, thus should not influence its stability (Figure 78 and Figure 80).

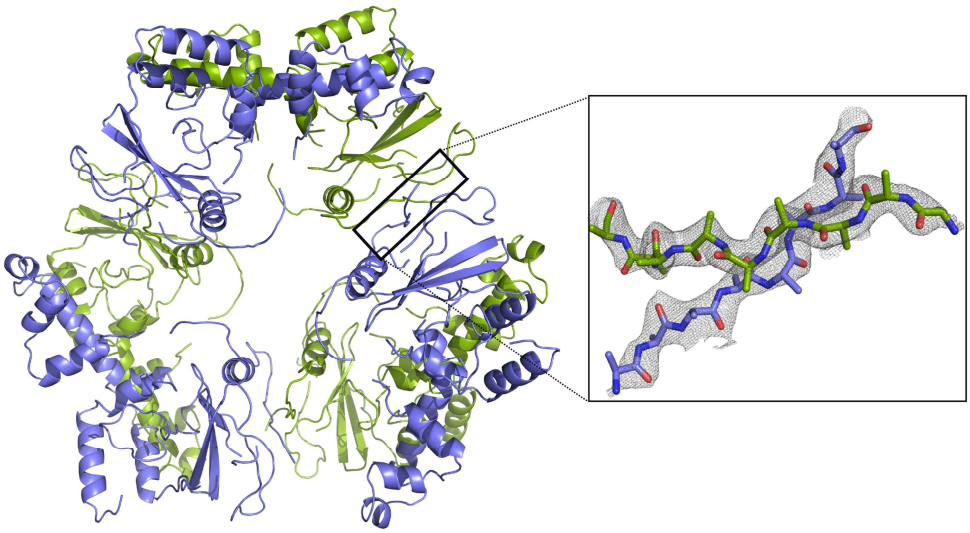

Figure 80: Close-up view of the hexamer interface of RovC. $2F_o$-F_c map (contoured at 1 σ) showing the β-strand conformation.

3) Mutation that disrupts the triskelion shaped hexamer (Set III) - The S219E/A220E double mutation will introduce six negative charges in the center of the triskelion assembly which should destabilize this hexamer. In contrast, these mutations are solvent exposed in the ring shaped hexamer and should thus not affect its stability (Figure 78 and Figure 81).

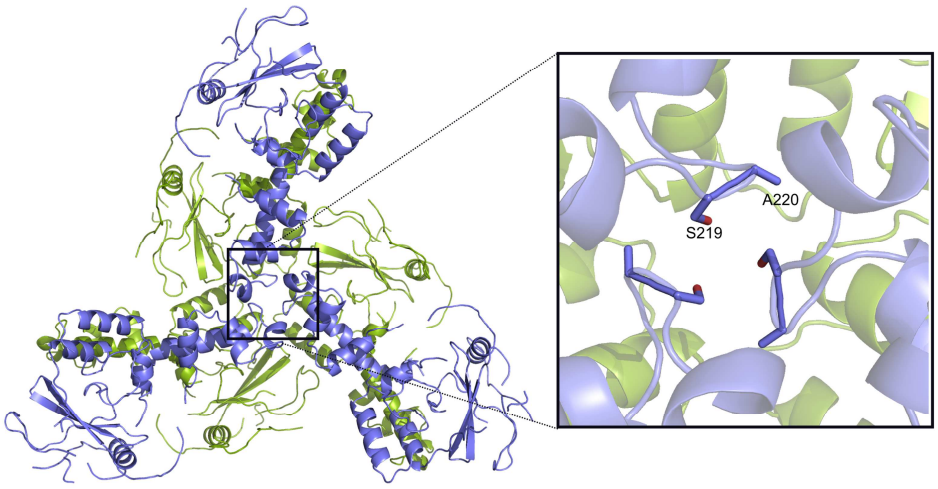

Figure 81: Close-up view of residues that were mutated in the triskelion shaped assembly.
Residues that were mutated (Set III) are labeled (only in one molecule) and shown as sticks with oxygen atoms in red.

Cloning, expression and purification of mutants

The above mentioned variants were cloned and mutations were verified by sequencing. Out of five different mutants tested for expression, only S219E/A220E and I150P resulted in soluble protein (Table 30).

Table 30: Overview on the solubility of different variants.

RovC variant	Hypothetical effect of the mutant	Solubility
Set I mutants		
A237E	Should disrupt the dimeric interface	Not soluble
G242E	Should disrupt the dimeric interface	Not soluble
Set II mutants		
I150P	Should disrupt the ring shaped hexamer	Soluble
I150P/Y151P	Should disrupt the ring shaped hexamer	Not soluble
Set III mutant		
S219E/A220E	Should disrupt the triskelion shaped hexamer	Soluble

Both mutants were expressed and purified in large scale (Figure 82). Expression and purification conditions were the same as described for the wild type protein.

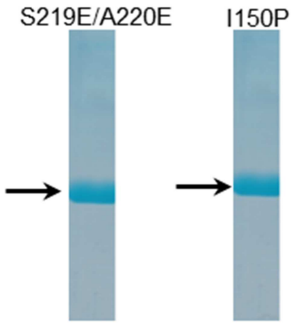

Figure 82: SDS gel showing purity of RovC mutants. Coomassie stained gel showing purified RovC variants. Bands corresponding to RovC variants are marked with the black arrow.

4.4.4.5 MALS with RovC variants

To further analyze the effect of mutations on the oligomeric state, MALS was performed. Molecular mass calculation revealed that S219E/A220E mutant forms a stable hexamer (calculated M.W. - 169 kDa) (Figure 83, shown in red). This result further validated the ring shaped model of RovC (Figure 78). For I150P mutant, five peaks were observed on the chromatogram (Figure 83, shown in green).

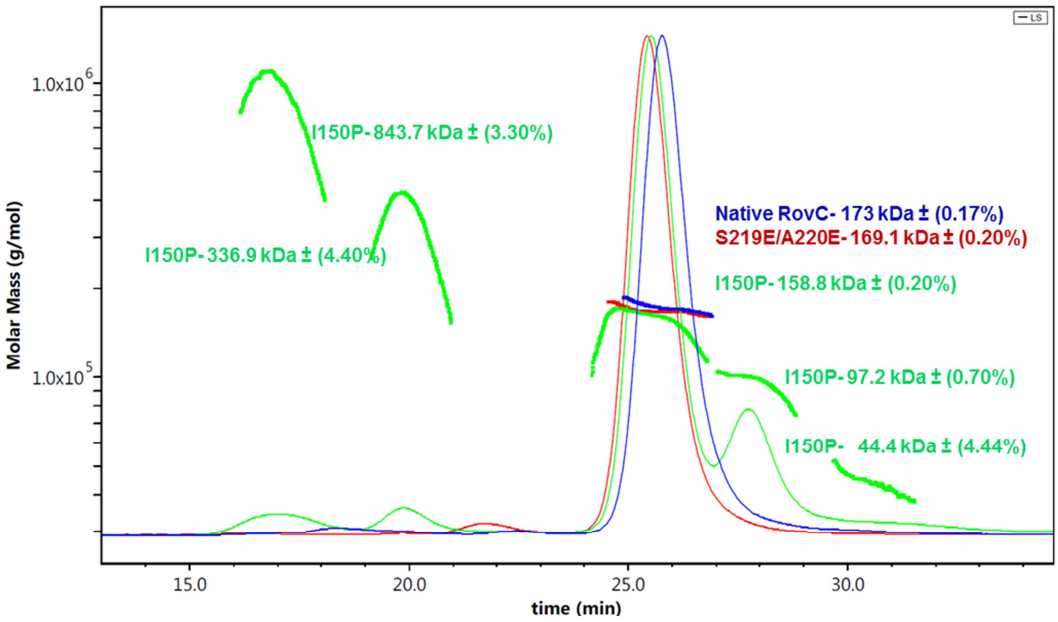

Figure 83: Light scattering chromatogram of RovC variants. Normalized SEC-MALS profile of native RovC (in blue), S219E/A220E mutant (red) and I150P mutant (green).

Molecular mass calculation revealed that 50% of the species form hexamers while the remaining form lower and higher order oligomers for which the exact oligomeric state could not be calculated with confidence. These results suggest that the I150P mutant, despite of being soluble, destabilizes the hexamer and is partially prone to unspecific, higher order oligomerization. The solubilty and MALS data, therefore strongly support the model of a hexameric ring-shaped structure of RovC in solution.

4.4.5 RovC interaction with T6SS4 DNA

4.4.5.1 Microscale thermophoresis

Vanessa Knittel (Prof. Petra Dersch's group) could show with electrophoretic mobility shift assays and DNAse footprinting assays that RovC binds to the promoter region of T6SS-4 (minimum binding fragment-38 bp). To further validate the interaction between purified RovC and Cy5-labeled T6SS4 DNA, microscale thermophoresis was used. In MST, instead of 38 bp, 42 bp fragment was used to avoid the influence of Cy5 on binding. Cy5-labeled DNA fragment was titrated with increasing concentrations of RovC in 100 mM TRIS pH 8, 500 mM NaCl, 5 mM DTT and 0.05%tween. Obtained results confirmed the binding of RovC to the 42 bp fragment with an experimentally determined affinity constant (K_d) in the range of 11.1 ± 3.8 µM. Moreover, DNA fragments of different lengths were used to find the minimal binding region. On reducing the length of oligonucleotides affinity decreased, which indicated that RovC binds to the complete 38 bp fragment (Figure 84).

A

5'gtagggataaacttattcgcagatttttttcacccttcatacattta 3'

gtagggataaacttattcgcagatttttttcacccttcatacattta	42-mer	K_d- 11.1 ± 3.8 µM
aaacttattcgcagatttttttcacccttca	30-mer	K_d- 92.2 ± 30.9 µM
ttattcgcagatttttttcaccc	22-mer	K_d- 406.7 ± 126.2 µM
tattcgcagatttttttcacccttca	25-mer_A	K_d- 331.7 ± 117.2 µM
aaacttattcgcagatttttttcacc	25-mer_B	curve cannot be fitted

B

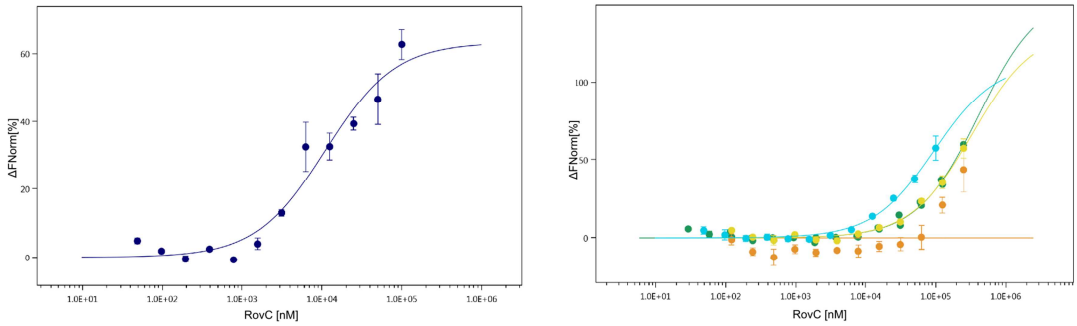

Figure 84: Interaction of RovC with T6SS4 DNA. A) RovC interaction with different fragments of Cy5 labeled DNA. B) MST curves showing the interaction of RovC with different DNA fragments, color scheme as in (A). Data shown is the average of duplicate experiments.

4.4.5.2 DNA binding region on RovC

In order to identify and narrow down the DNA-binding region on RovC, the electrostatic potential and conservation has been plotted on the surface of the RovC hexamer. An extended basic patch might be involved in interaction with the negatively charged backbone and conserved residues might be involved in specific contacts with the target DNA.

1. Electrostatic surface potential of RovC-

   The surface potential of hexameric RovC was calculated in pymol using the generate vacuum electrostatics function (Baker et al., 2001). The analysis revealed one basic patch (per dimer) on the outer surface of the ring mainly formed by residues 225-240 present on the helix-turn-helix motif, which represents the potential DNA binding site (Figure 85A).

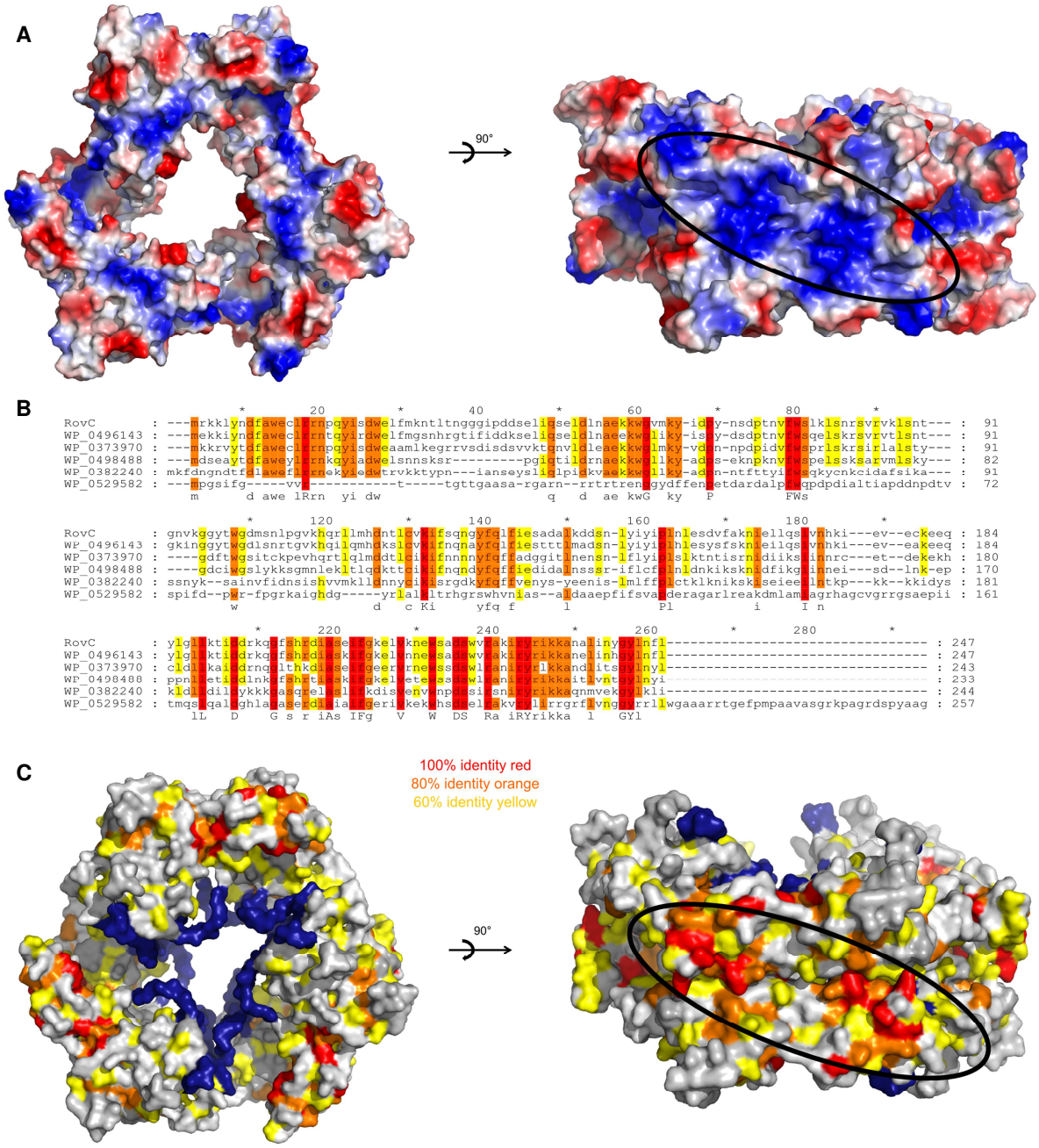

Figure 85: Potential DNA binding site on RovC. A) Electropotential surface analysis of RovC shown in two orientations. Potentials between -74.194 kT/e (red) and + 74.194 kT/e (blue) are color-coded. B) Alignment of RovC related proteins (red: 100% conserved amino acids; orange: 80%; yellow: 60%) (WP_049614373-hypothetical protein from *Y. pekkanenii*; WP_037397067- hypothetical protein from *Serratia*; WP_049848866 hypothetical protein from *Trabulsiella odontotermitis*; WP_038224070-hypothetical protein from *Xenorhabdus bovienii*; WP_052958275-hypothetical protein from *Mesorhizobium* sp. LC103) (Alignment generated with GeneDoc, (Nicholas et al., 1996)). C) Conserved regions on the surface of RovC shown in two orientations. Color scheme as in (B). The dark blue region corresponds to the poly-ala stretch with ill-defined density in the structure, to which a conservation could not be mapped onto.

2. Sequence conservation of RovC-

Analysis based on the primary sequence identified six hypothetical proteins in the NCBI database that share 45-80% sequence identity with RovC (Figure 85B). These proteins are from different Gram-negative bacteria. As the sequence conservation is very high, these proteins are likely to adopt a similar fold and possess DNA binding function as RovC. The sequence conservation of these proteins was mapped onto the structure of RovC. As shown in (Figure 85C), residues 219-240 in the C-terminal domain, a region that partially overlaps with the basic patch, are highly conserved while the N-terminal domain is less strongly conserved.

Sulfate ions in the crystal structure

During the structure refinement, pieces of spherical electron density were observed that likely represent sulfate ions because of the crystallization condition containing $MgSO_4$ as precipitant. All of the sulfate ions lie on the possible DNA binding site and might mimic DNA backbone phosphates (Figure 86).

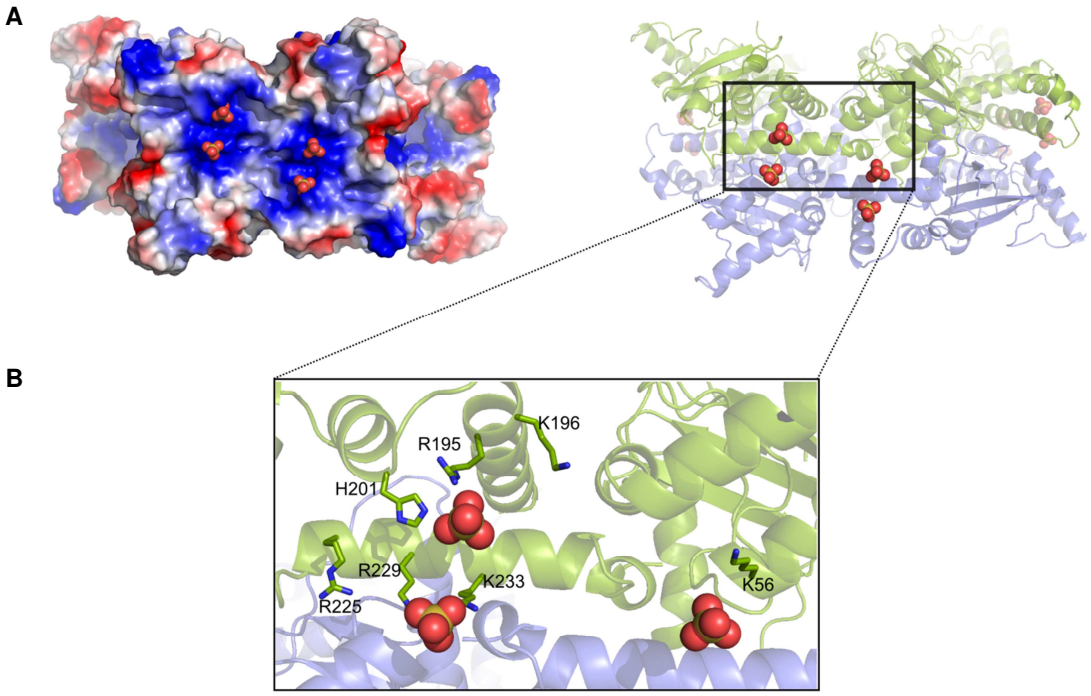

Figure 86: RovC structure with bound sulfate. A) The electrostatic potential map of RovC showing presence of sulfate ions on the positively charged patch. Sulfate ions are shown as spheres with sulfur atoms as yellow and oxygen atoms as red. Potentials between -74.194 kT/e (red) and + 74.194 kT/e (blue) are color-coded. B) Close-up view of the positively charged residues that are close to sulfate ions in the structure. Residues are labeled and shown as sticks with oxygen atoms in red and nitrogen atoms in blue.

Mutational analysis of DNA binding site

To precisely analyze the region for DNA binding, site directed mutagenesis of the basic and solvent exposed residues (R225, R229 and K56) located on this patch was performed. These residues are conserved in RovC sequence as shown in Figure 85B. Apart from the above mentioned residues, two non-conserved lysines (K211/K215) present on the helix-turn-helix motif were also analyzed for DNA binding (Figure 87).

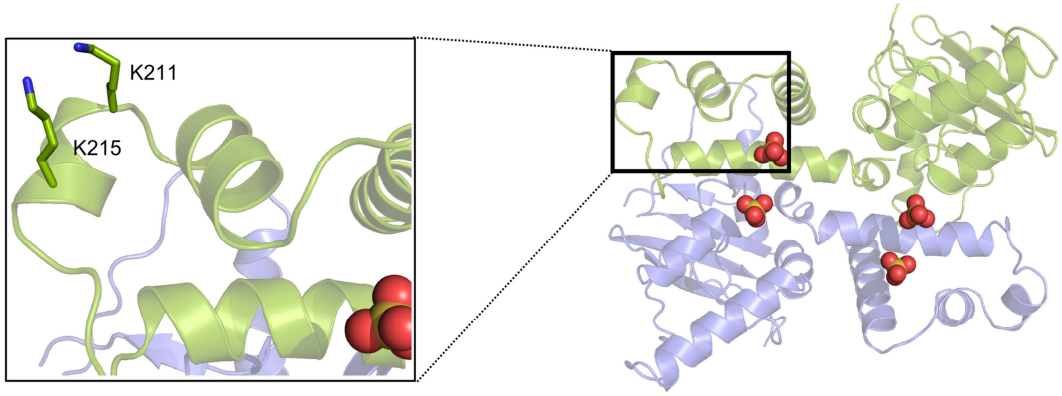

Figure 87: Close-up view of residues that were mutated on the helix-turn-helix motif. Residues that were mutated are labeled and shown as sticks with oxygen atoms in red. Sulfate ions are shown as spheres with sulfur atoms as yellow and oxygen atoms as red.

These residues were mutated to Glutamate. Mutants were cloned, expressed, and purified (Figure 88) as described for the wild type protein.

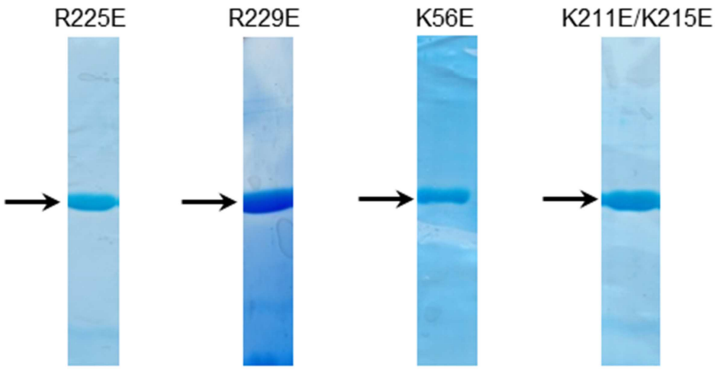

Figure 88: SDS gel showing purity of RovC mutants. Coomassie stained gel showing purified RovC mutants. Bands corresponding to RovC variants are marked with the black arrow.

MALS with R225E, R229E, K56E and K211E/K215E variants

Prior to examining the DNA binding activity of mutants, MALS was performed to ensure that the mutations did not affect the hexameric state of protein. As shown in Figure 89, all mutants form hexamers in solution.

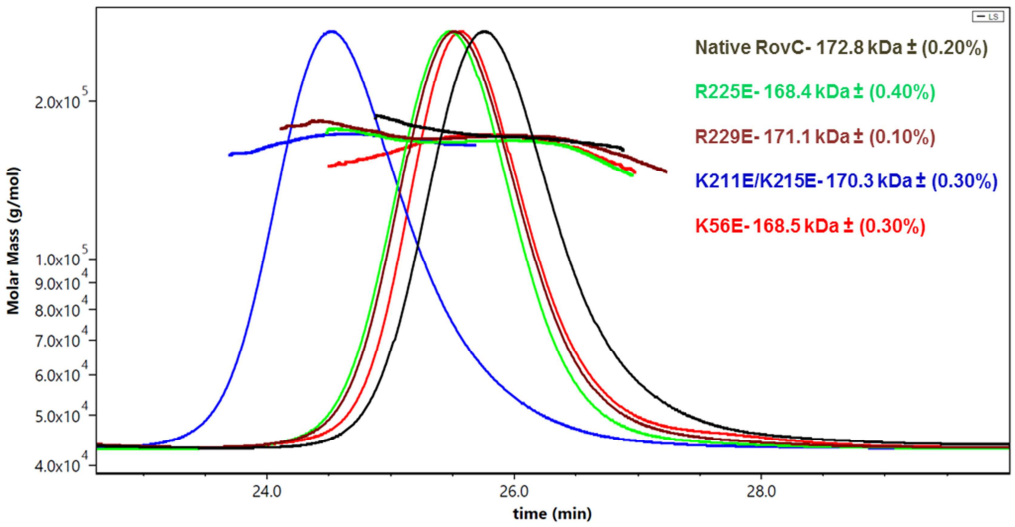

Figure 89: Light scattering chromatogram of RovC variants. Normalized SEC-MALS profile of native RovC (in brown), R225E mutant (green), R229E mutant (dark red), K211E/K215E mutant (blue) and K56E mutant (red).

EMSA

Mutants that form stable hexamers in solution (R225E, R229E, K56E, K211E/K215E and S219E/A220E) were tested for DNA binding in EMSA by Vanessa Knittel. R225E, R229E and S219E/A220E completely abolished DNA binding activity while K56E and K211E/K215E mutants had no influence on DNA binding suggesting that the DNA binding region lies on the conserved basic patch of RovC (discussed in detail in 5.4.3.1).

4.4.5.4 Co-crystallization of RovC and T6SS4 DNA

As the mutagenesis studies can identify basic residues involved in interacting with the DNA backbone, residues involved in recognition of a specific DNA sequence cannot easily be identified by mutagenesis studies. Therefore, co-crystallization of RovC with the 38 bp DNA binding fragment was performed using different commercially available screens at 20°C and 4°C. RovC and DNA were mixed in a molar ratio of 1:1.2 prior to crystallization set-ups. However, no co-crystals were obtained.

As the length of DNA plays a significant role in crystallization of protein-DNA complex (Pakotiprapha and Jeruzalmi, 2014), therefore, blunt ended DNA fragments of various lengths were used (Table 31). But this approach also did not result in co-crystal formation.

Table 31: DNA fragments used for co-crystallization with RovC.

Name	**Sequence 5'→3'**
38 mer	ggataaacttattcgcagatttttcacccttcataca
36 mer	gataaacttattcgcagatttttcacccttcatac
34 mer	ataaacttattcgcagatttttcacccttcata
32 mer	taaacttattcgcagatttttcacccttcat

5 Discussion

5.1 InvD from YPIII strain of *Y. pseudotuberculosis*

Y. pseudotuberculosis possess several multifunctional adhesins on their surface that are not only needed by bacteria for binding to the host cell surface but are also required for efficient internalization and colonization. InvD has been recently identified in the YPIII strain of *Y. pseudotuberculosis* (Pisano et al., 2012), but the information on the structure and function has remained elusive. Thus, in this study, InvD structure was solved, its host receptor was identified and characterized by biochemical and biophysical experiments.

5.1.1 The Big12 and Big13 domains

InvD consists of 13 Big domains out of which Big1-12 are highly repetitive in sequence while Big13 is significantly different from Big1-12 (Figure 24). Despite the low sequence similarity, Big13 also adopts a β-sandwich fold, similar to Big12. Big domains are identified in a variety of other bacterial surface proteins such as in Lig proteins from *Leptospira interrogans* (Raman et al., 2010), SiiE from *Salmonella enterica* (Griessl et al., 2013) and InvA from *Y. pseudotuberculosis* (Hamburger, 1999). These domains have a variety of different functions as seen for Big domains of SiiE that target carbohydrates present on the host cell surface (Barlag and Hensel, 2015) or as a linker, to span a certain distance between the outer membrane anchor and the C-terminal adhesion domain, as reported for the D1-D3 domains of InvA (Hamburger, 1999). Big12 domain of InvD adopts an I1-set of IgSF fold (Figure 25). As Big1-11 share high sequence similarity with Big12, they are likely to adopt a similar fold. I-set of IgSF fold is often seen in other bacterial domains that act as linker (Wang, 2013). In contrast, the difference in the sequence of Big13 domain suggests a special role of this domain in bacterial pathogenicity along with the adhesion domain. This was seen in case of InvA where D4 domain interacts with the head domain (D5) to form a superdomain necessary to bind β_1-integrins (Hamburger, 1999). Thus, based on the limited proteolysis data of InvD1737/InvD1640 constructs, SAXS and structural data, it can be concluded that Big13 and AD form a functional superdomain with low conformational flexibility, while Big1-12 act as spacer modules.

5.1.2 The adhesion domain of InvD

Unlike InvB and InvC adhesion domains, InvD AD shares no sequence similarity with the C-type lectin domain of InvA. The core structure of the InvD AD adopts a β-sandwich fold and resembles the C2-set fold with extensive variations and insertions. Comparison of InvD-AD topology with Big12/13 suggests that *Yersinia* has evolved the AD from its Big domain to carry out a specific function (Figure 30).

Sequence analysis of InvD-AD using NCBI BLAST-p identified four structurally uncharacterized proteins that share more than 65% sequence identity (similarity 76-97%) with the AD of InvD (Figure 90A). All these proteins belong to different strains of the species *Y. pseudotuberculosis*. As considerable sequence similarity exists between InvD AD and the above mentioned proteins, these proteins are likely to adopt a similar fold, and thus probably fulfill a similar function as that of InvD YPIII adhesion domain. The sequence conservation of these proteins was mapped onto the three-dimensional structure of InvD to identify potential functional patches with high conservation rate (Figure 90B).

Two major, highly conserved patches were identified at the top and the bottom of InvD structure. The first one is a large surface generated by the C'CFG sheet extending to the opposite sheet on strand A and the short antiparallel β-strands insertion between strands B and C. The second patch is located at the tip of the molecule and comprises residues of the α-helix and the C*/C'* β-turn. The two highly conserved patches are separated by a large surface forming a deep cleft in the protein with lower conservation, extending from strands C*/C'* to strands E, B and partially A (shown with black dotted lines in Figure 90B).

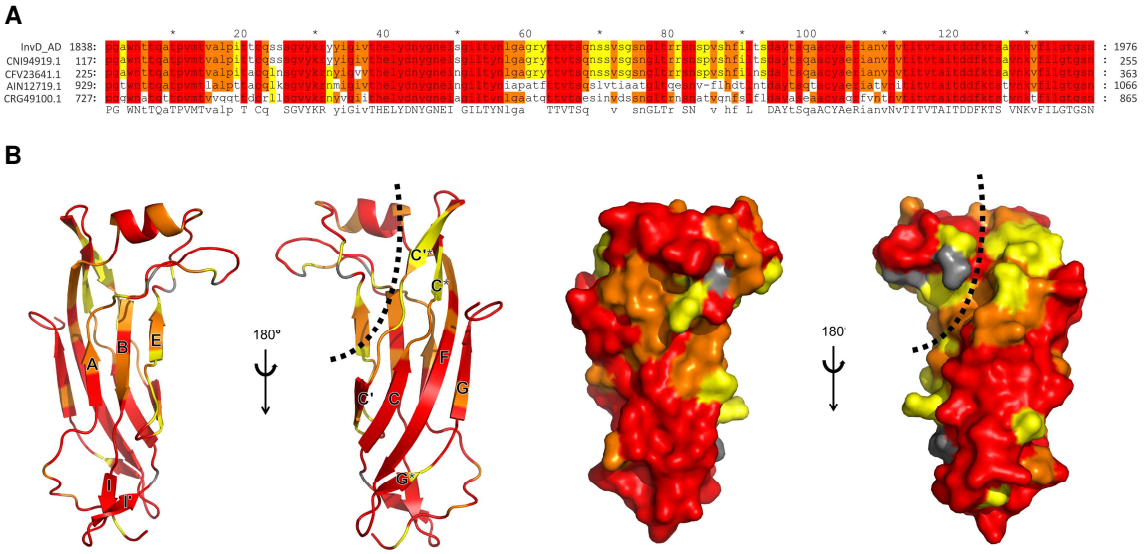

Figure 90: Conservation patches on InvD AD. A) Alignment of InvD related proteins (red: 100% conserved amino acids; orange: 80% and yellow: 60%). CNI94919.1, CFV23641.1 and AIN12719.1 belong to different strains of *Y. pseudotuberculosis*; CRG49100.1 belongs to the species *Y. wautersii.* (Alignment was generated with Genedoc, (Nicholas et al., 1996)). B) Conserved regions on the structure of InvD AD as cartoon representation (left) and surface representation (right) shown in two orientations. Color scheme as in (A). The cleft is indicated with dotted lines.

These proteins are likely to provide a similar function as that of InvD, albeit possibly targeting different subtypes of the receptor of the same or different hosts, suggesting a broader use of this strategy by different zoonotic strains of the *Yersinia* genus.

In this deep cleft, some additional electron density was identified (Figure 91). Since InvD used for crystallization still contains the hexahistidine-tag, the tag peptide from a symmetry-related molecule might occupy the position in the cleft resulting in the additional density, to which a sequence could however not be assigned to due to poor quality of the density.

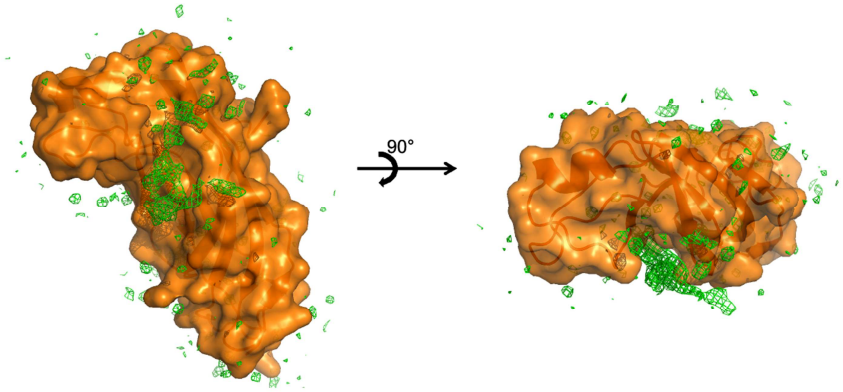

Figure 91: Electron density in the cleft of InvD-AD. Surface representation of InvD AD with the electron density of the tag (F_o-F_c map contoured at 2.5 σ) in the cleft region shown in two orientations.

5.1.3 InvD interaction with immunoglobulins

Structural similarity search using InvD AD did not result in any statistically significant match, which hampered the assignment of function to InvD AD on the basis of its structural features (Figure 31). Therefore, pull down experiments using homogenized extract of mice organs were performed which identified antibodies as the potential interaction partner of InvD. Further, *in vitro* panning and interaction assays confirmed that InvD AD specifically targets the Fab part of immunoglobulins of the VH3/VK1 subfamily combination with nanomolar to low micromolar affinity.

These results strongly indicate that InvD-Fab interaction is not a classical antigen-antibody interaction since,

i) antibodies isolated after 3rd of panning have common variable heavy (VH) and variable light (VL) chain subtypes, but different Complementarity Determining Regions (CDRs),

ii) sequence alignments of VH and VL CDRs indicate broad diversity of CDR H3. CDR H3 are largely responsible for antigen specificity as these represent unique loops that arise from the hypervariable nature of the VDJ region (Figure 92),

iii) and binding of InvD to specific subtype of the antibody is not dependent on prior immune exposure since „unselected libraries“ were used for panning.

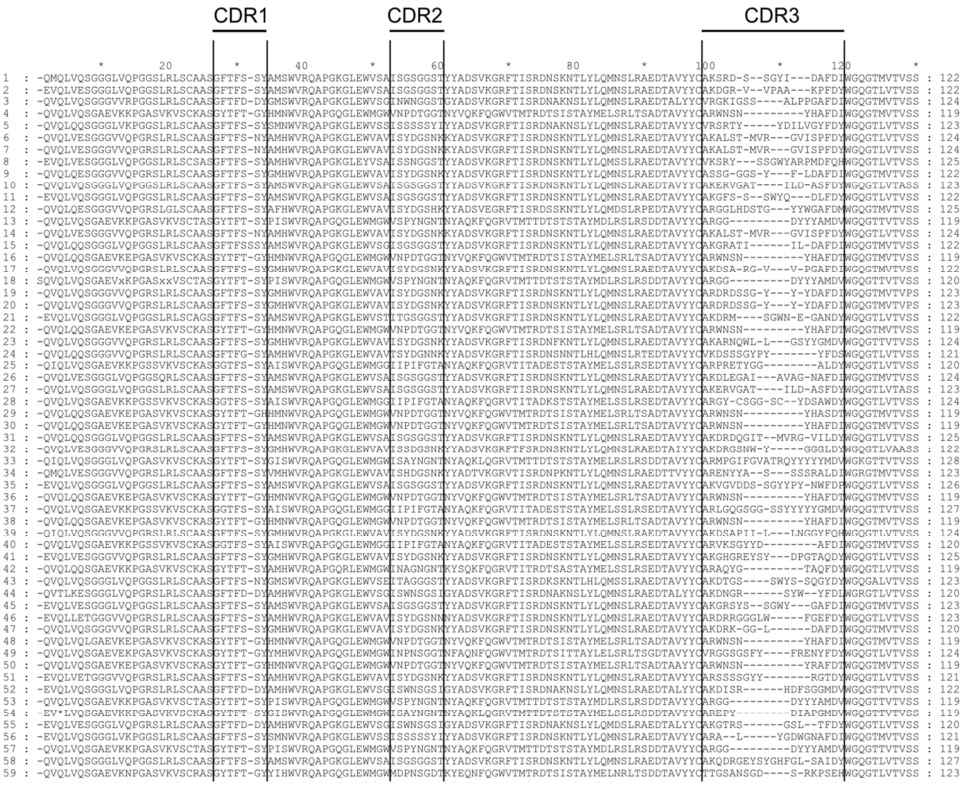

Figure 92: Sequence alignment of antibodies isolated after 3rd round of panning. Alignment of 59 sequenced VH fragments (from HAL 9+10 libraries) indicates broad diversity in CDR 3 region. Alignment is generated with GeneDoc (Nicholas et al., 1996).

Previous studies have demonstrated that bacterial and viral proteins that target specific subtype of the Fab region of immunoglobulins act as B-cell superantigens. Examples are staphylococcal Protein A (SpA) from *Staphylococcus aureus* and gp120 from human immunodeficiency virus-1 (HIV) (Townsley-Fuchs et al., 1997), which both target the VH3 subfamily of Fab (Sasso et al., 1989) and protein L from *Peptostreptococcus magnus*, which binds to VK1, VK3 and VK4 subfamilies (Björck, 1988). These proteins are known to target not only the soluble antibodies but also the membrane bound antibodies exposed on B cells (Silverman and Goodyear, 2006) and can cause B–cell apoptosis, proliferation or overactivation of the classical complement pathway (Gjörloff Wingren et al., 2002; Kozlowski et al., 1996; Ruffet et al., 1994; Townsley-Fuchs et al., 1997).

With the high specificity binding to the VH3/VK1 antibody subfamily combination (Figure 40 and Figure 41) and binding to the B-cells surface (Figure 38), InvD shares these properties with bacterial and viral proteins that act as B-cell superantigens, however there are some differences.

In contrast to classical superantigens, that target more than 5% of the B cell population by binding to either VL or VH chain domain of the membrane bound antibody (Silverman and Goodyear, 2006), InvD targets a combination of VL and VH, and binds to ~0.5% of murine B-

cell population. Therefore, future experiments are required to understand the biological significance of InvD-immunoglobulins interaction.

In the present study, the roles of InvD to interact with VH3/VK1 fragment of the antibody and B cells are established. But, how and at which stage of the infection process InvD targets antibodies and/or B cells remain interesting open questions to be answered by future experiments.

5.2 InvE from IP31758 strain of *Y. pseudotuberculosis*

Y. pseudotuberculosis produces different adhesins which are not only essential for bacterial uptake but are also needed for efficient colonization in the host. Strain IP31758 of *Y. pseudotuberculosis* is responsible for Far East scarlet–like fever and also leads to toxic shock syndrome, which is not commonly seen for *pseudotuberculosis* infections (Eppinger et al., 2007). A novel Invasin, InvE has been identified in this strain. So far, there has been no report on the structural and functional characterization of InvE. In this study, the three dimensional structure of InvE adhesion domain along with the two BIg domains was determined by X-ray crystallography and its structural characteristics are discussed.

5.2.1 The BIg20 and BIg21 domains

InvE consists of 21 BIg domains, out of which BIg1-20 share high sequence similarity while BIg21 is significantly different from BIg1-20 (Figure 52). Despite their weak sequence similarity, both BIg20 and BIg21 adopt an IgSF fold. BIg20 belongs to the I1–set while BIg21 is categorized in the I2 set of IgSF fold (Figure 53). I-set of Ig-like fold generally acts as a linker to span a certain distance and present the ligand binding domain on the cell surface (Wang, 2013). Comparison of BIg21 with other Ig-like domains of InvA and InvE reveals a unique conformation (Figure 56), important for its interaction with the AD. Structural and sequence difference in the last Ig like domain suggests its role in bacterial pathogenicity along with the AD, as seen with InvA where D4-D5 domain interface spans significantly larger area required to form a superdomain necessary to bind β1-integrins (Hamburger et al., 1999). Based on the limited proteolysis, SAXS and structural data, it is evident that in the rod shaped structure of InvE, BIg21 and AD form a functional superdomain with low conformational flexibility, while BIg1-20 act as spacer modules.

5.2.2 The adhesion domain

Structure based search of InvE adhesion domain using PDBefold resulted in the best fit with the CTLDs of Intimin from EPEC/EHEC and InvA from *Y. pseudotuberculosis.*

Both Intimin and InvA are ~900 aa protein, consisting of the N-terminal β–barrel domain, followed by Ig like domains and the CTLD. Crystal structure of InvA reveals four Ig like domains (D1-D4) and the C-terminal domain (D5) while Intimin comprises three Ig like domains (D0-D2) and the C-terminal domain (D3) (Hamburger et al., 1999; Luo et al., 2000; Yi et al., 2010). During the infection process, Intimin targets the translocated intimin receptor (Tir), which is exported by the bacteria and integrated into the host cell plasma membrane (Kenny et al., 1997), whereas InvA targets β_1-integrins on the host cell membrane, mediating efficient cellular uptake of the pathogen (Alrutz et al., 2001; Isberg et al., 1987).

InvE AD shares high structural similarity (Z score>8.5) with the adhesion domain of InvA and Intimin (Figure 93A, B) despite less than 17% sequence identity (Figure 93C). When InvE-CTLD is superimposed on the adhesion domain of InvA (PDB ID-1cwv:A) and the adhesion domain of Intimin (PDB ID-3ncx:A), the rmsd is 2.5 Å and 2.2 Å, respectively.

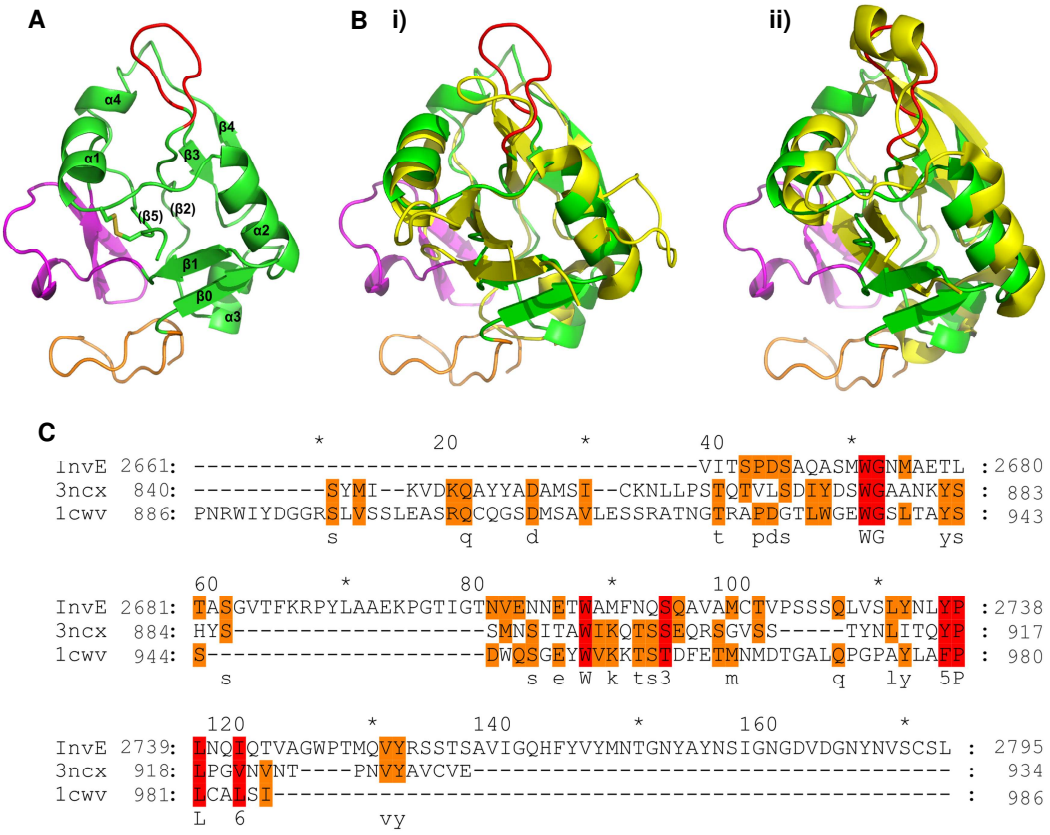

Figure 93: Structural and sequence alignment of InvE AD. A) Cartoon representation of InvE AD with β-strands and α-helices labeled (color scheme as in Figure 59). B) Structural comparison of InvE AD with other structurally similar proteins. Superposition of InvE AD (color scheme as in Figure 59) with (in yellow) i) Intimin adherence protein (PDB ID-3ncx:A) and ii) InvA (PDB ID- 1cwv:A). C) Sequence alignment of InvE AD (InvE) with Intimin adherence protein (3ncx) and InvA (1cwv). (100% conserved amino acids are marked in red and 80% in orange). Alignment is generated with GeneDoc (Nicholas et al., 1996).

However, the major structural differences among them occur at the N-terminus which includes: presence of strand β0, wedge domain and adaptor ring in InvE AD, whereas InvA and Intimin lack these elements (Figure 93A, B). While the AR in InvE appears to stabilize the overall rod-shaped structure and link the BIg21 to AD as shown in Figure 57A, the function of the WD is at this stage less clear.

Another difference among them is the relative orientation of the adhesion domains to the residual part of the protein. InvA and Intimin share a similar orientation while the orientation of InvE AD is different as shown in Figure 94.

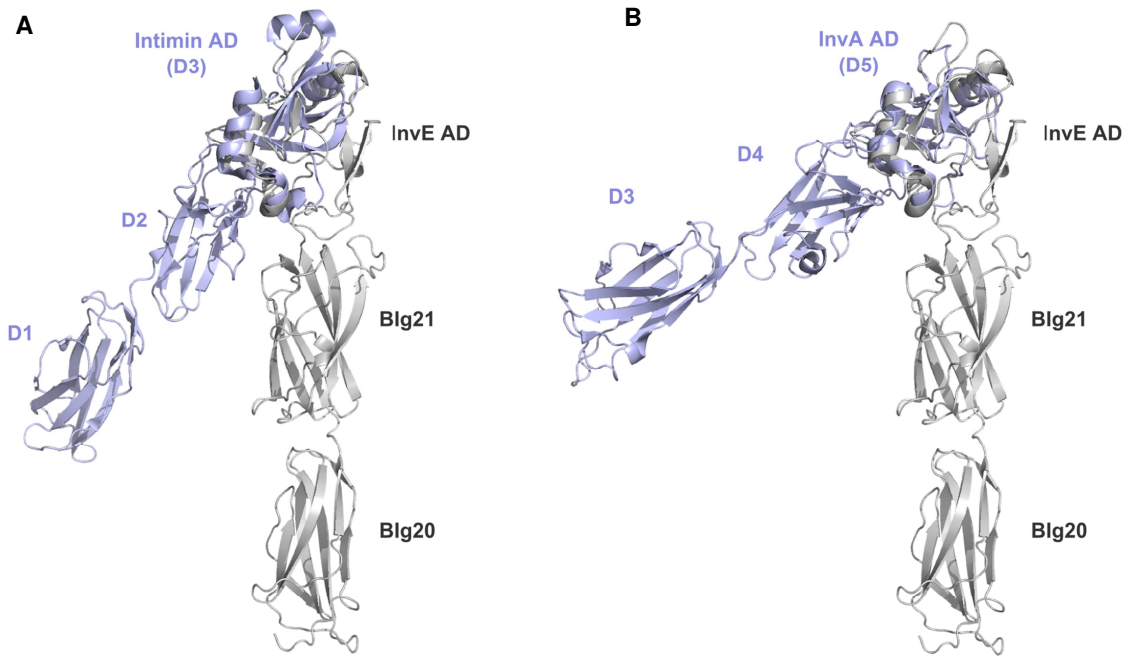

Figure 94: Orientation of InvE AD as compared to InvA and Intimin. Structural overlay of InvE (Big20/Big21/InvE AD, in gray) with (in light blue) A) Intimin (D1/D2/Intimin AD) and B) InvA (D3/D4/InvA AD).

As described above, the CTLDs of both InvA and Intimin promote bacterial adherence.

1. In InvA, the integrin binding site is formed by polar and charged residues located on helix α1 and the loop followed by it (Hamburger et al., 1999) (Figure 95B). While InvE AD lacks cluster of residues required for integrin binding and contains mainly hydrophobic residues (Figure 95C).

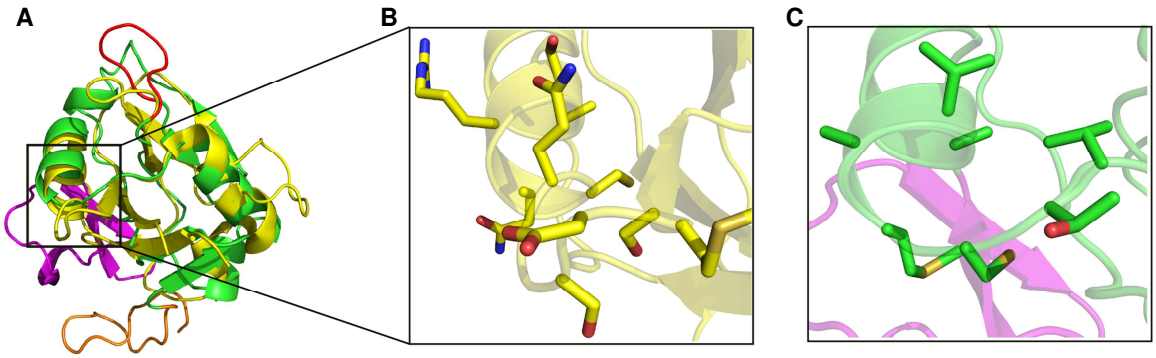

Figure 95: Close-up view of the Integrin binding site in InvA and its comparison with InvE.
A) Structural overlay of InvA (PDB ID-1CWV:A) and InvE adhesion domain (color scheme as shown in Figure 93A). B) Residues involved in binding integrin are shown on the structure of InvA and C) the corresponding residues in InvE AD are shown. These residues are shown as sticks with sulfur atoms in yellow, oxygen atoms in red and nitrogen atoms in blue.

2. Structural superposition of EPEC Intimin-Tir structure (PDB ID-1f02) with InvE AD shows that when Intimin is replaced by InvE in the complex structure, steric clashes arise (marked with a black dotted circle in Figure 96C). Furthermore, InvE lacks the short helix (marked with a black arrow in Figure 96B) that is involved in Tir binding in Intimin (Luo et al., 2000) (Figure 96B, C).

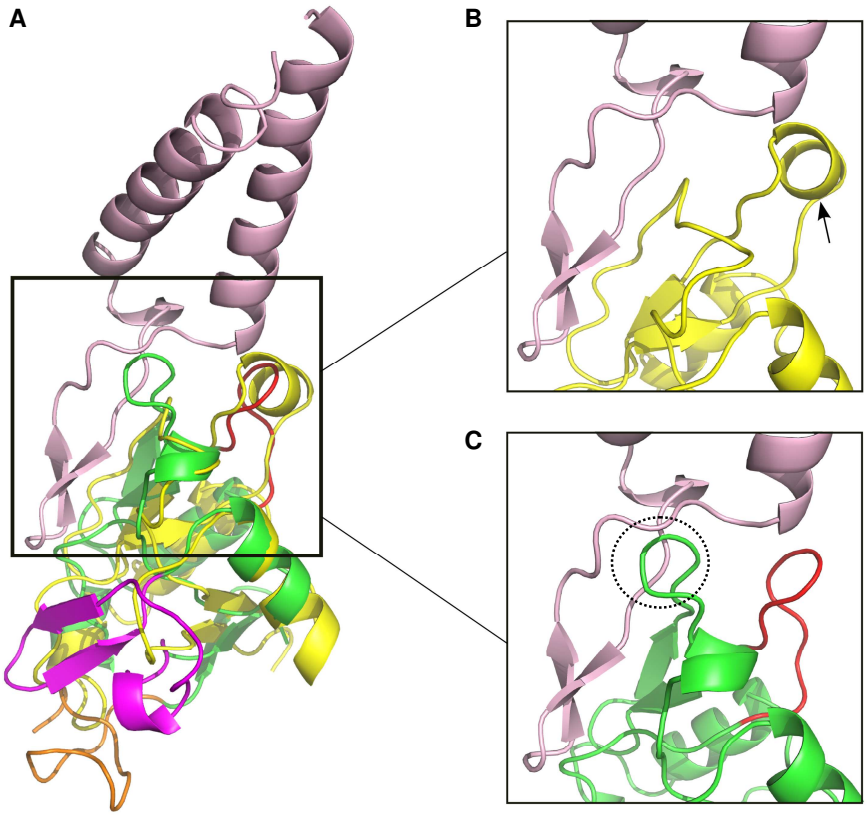

Figure 96: Close-up view of Intimin-Tir interaction interface and its comparison with InvE.
A) Structural overlay of EPEC Intimin-Tir structure (PDB ID-1f02) (Intimin in yellow and Tir in light pink) with InvE AD (color scheme as in Figure 93A). B) Close-up view of Intimin-Tir and C) InvE AD-Tir interaction interface. Helix is marked with a black arrow in B. Black dotted circle in panel C represents the steric clashes in between InvE and Tir.

Thus, despite the overall structural similarity of the adhesion domain of InvE to InvA and Intimin, it appears that the AD of InvE has evolved to target a distinct receptor.

InvE AD lacks the cluster of residues required for binding to Ca^{2+} and also motifs such as EPN, QPD and WND that are responsible for the specificity of sugar recognition (Zelensky and Gready, 2005).

The identification of CTLD in InvE suggests its role in targeting a specific receptor on the host cell. The C-type lectin fold is at the core of many proteins with a wide range of functions including adhesion, signaling and cell-cell recognition (Zelensky and Gready, 2005), which hampers the assignment of function to InvE AD on the basis of its structural features.

5.3 YadA$_{pstb}$

YadA$_{pstb}$ and YadA$_{ent}$ belong to the trimeric autotransporter family and share high sequence homology among themselves, however they differ in their pathogenicity. The difference in their virulence behavior is due to the presence of an additional 31 aa in YadA$_{pstb}$ head region, that constitutes the ‘’uptake domain’’. Deletion of the 31 aa motif abolishes fibronectin binding and the molecule preferentially binds to collagen (Heise and Dersch, 2006). But the structural basis for this functional transition is not known.

The structure of the head domain of YadA$_{pstb}$ is solved in this work and compared with YadA$_{ent}$. The overall architecture of both proteins is similar consisting of β-roll trimeric structure (Figure 63). Both have NSVAIG-x-S motif required for collagen binding (Nummelin et al., 2004). However, the 31 aa region in YadA$_{pstb}$ could not be completely built which indicates that this region is highly flexible. Nonetheless, the insertion was identified to be located at the tip of trimeric structure (Figure 97) which might modify the surface of the tip region of YadA$_{pstb}$ in comparison to YadA$_{ent}$, which is likely to result in the changed binding specificity from collagen in YadA$_{ent}$ to fibronectin in YadA$_{pstb}$. Even though the crystal structure has been solved, with the current data, it is difficult to give a detailed answer to the question on target specificity.

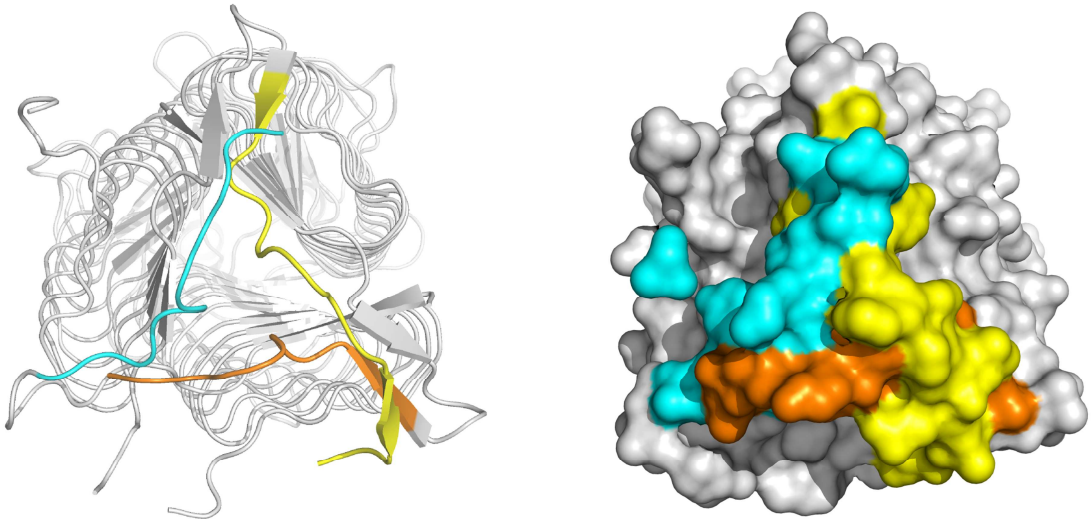

Figure 97: Top view onto the tip of YadA$_{pstb}$. Structured parts of the 31 aa insertion are shown in cyan, orange and yellow (left) on the cartoon model, (right) and surface representation.

5.4 RovC

Expression levels of various virulence factors in *Yersinia* are controlled by complex regulatory networks. The regulation is achieved at transcriptional, post-transcriptional and translational level by several proteins and RNAs (Erhardt and Dersch, 2015). One of these is RovC that regulates T6SS4 and Csr systems of *Yersinia.* Role of RovC in virulence has been established by the group of Prof. Petra Dersch (dissertation Stephanie Christine Seekircher, 2014/Vanessa Knittel, unpublished data), but the details on the structure of RovC and the regulatory mechanism were not known. In order to gain insights into the molecular mechanism of RovC, its structure was solved and further characterized using biophysical and biochemical methods.

5.4.1 Hexameric ring shaped structure of RovC

In this study, the crystal structure of RovC was determined to 2.3 Å resolution. RovC is composed of the N-terminal domain, linker and the C-terminal domain. Further, SEC-MALS data revealed that RovC assembles as a hexamer in solution. Although, the determined structure contained one molecule in the asymmetric unit, hexameric RovC was generated by crystallographic symmetry in which the three dimers are arranged to form a ring like structure. The ring shaped model was further validated by SAXS and mutational studies (S219/A220 and I150P mutants) using SEC-MALS.

RovC hexameric interface (Interface I as shown in Figure 98) forming the ring is formed solely by the N-terminal domain.

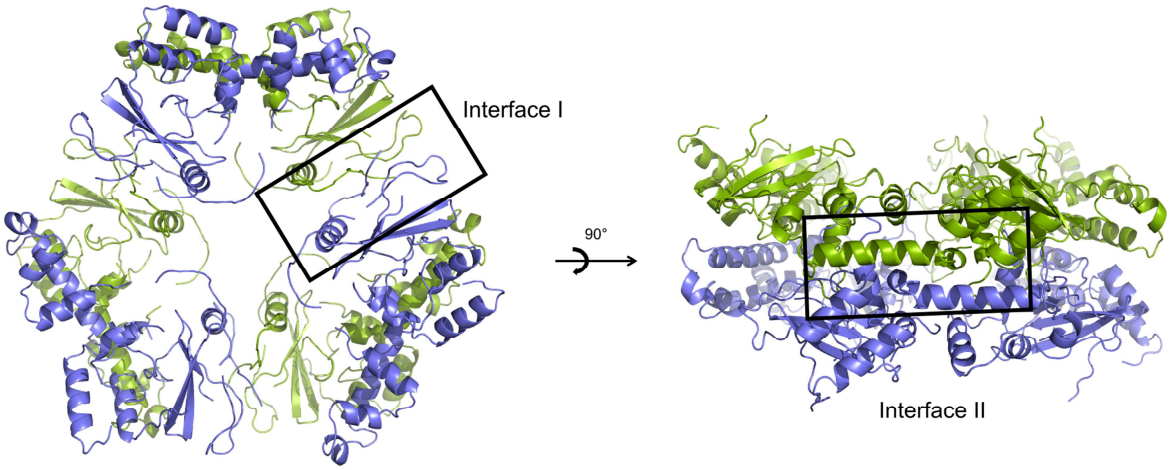

Figure 98: RovC hexameric and dimeric interface. The hexameric interface (interface I) (left) and the dimer interface (interface II) (right) are marked with a black box.

The interface and the linker region were partially built as poly-alanine chain into the electron density which precludes the possibility of making detailed interpretations regarding its structural organization and interface arrangement. This also indicates that the interface might allow for minor flexibility between different subunits.

RovC dimer interface (Interface II as shown in Figure 99) was calculated by PISA. The interface is formed by 12 N-terminal and 16 C-terminal residues (per monomer) comprising an interface area of 1142.5 Å^2 (formed by 14 hydrogen bonds and 2 salt bridges). Additionally, a network of hydrophobic interactions across the dimer interface were analyzed using the DIMPLOT module of LigPlot+ (Laskowski and Swindells, 2011) as shown in Figure 99.

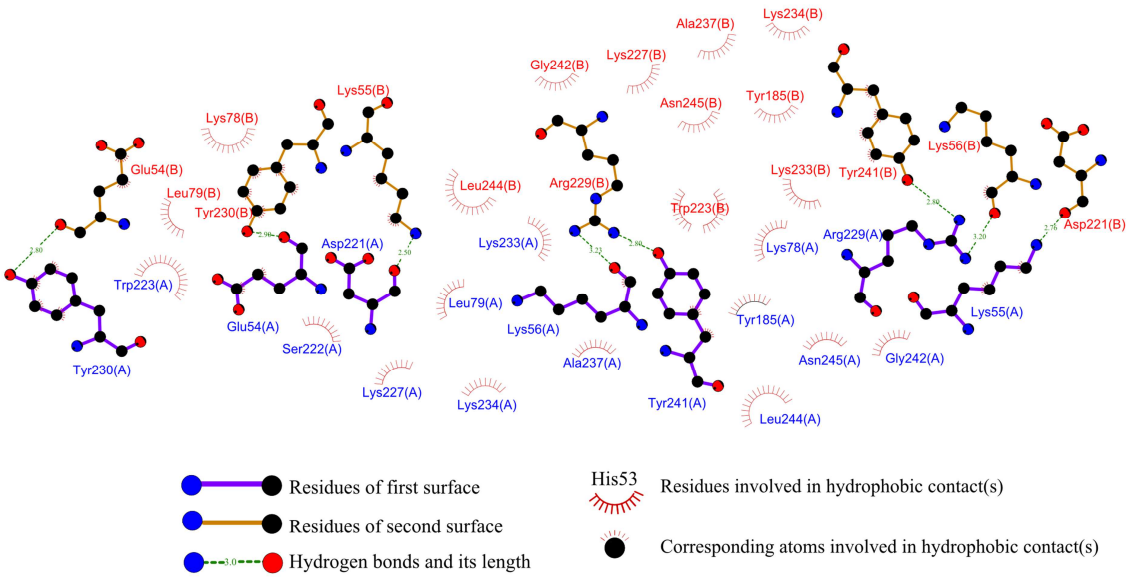

Figure 99: Dimplot analysis of the dimeric interface of RovC. Hydrogen bonds are shown by green dashed lines and spoked arcs represent hydrophobic interactions. Residues that belong to chain A are labeled in blue while chain B residues are in red.

All of the mutants with dimerization interface mutations resulted in insoluble protein. I150P mutant, which was designed to disrupt the ring formation, resulted in the formation of different order oligomers, while the double mutant I150P/Y151P resulted in insoluble protein. This indicates that the hexamer formation is critically important since any mutation designed to disrupt the dimeric or hexameric interface resulted in insoluble protein or unstable hexamer.

5.4.2 Structural similarity search with DALI

The N and C-terminal domains of RovC were separately submitted to the DALI server (Holm and Rosenstrom, 2010), as no structures were found that shared significant similarity with the full length RovC.

For the N–terminal domain, no significant match (Z score<2.8) was obtained with the top hits being lactoylglutathione lyase (PDB ID-2qh0:A), an enzyme that catalyzes the isomerization of the hemithioacetal (Thornalley, 2003) and Bleomycin acetyltransferase (PDB ID-2zw5:A), an enzyme that inactivates Bleomycin using acetyl-coenzymeA (Oda et al., 2010). Structural superposition of the N-terminal domain of RovC with these proteins however shows that there is no similarity in topology or fold with a high rmsd>4.5 Å, indicating that this domain adopts a unique fold (Figure 100).

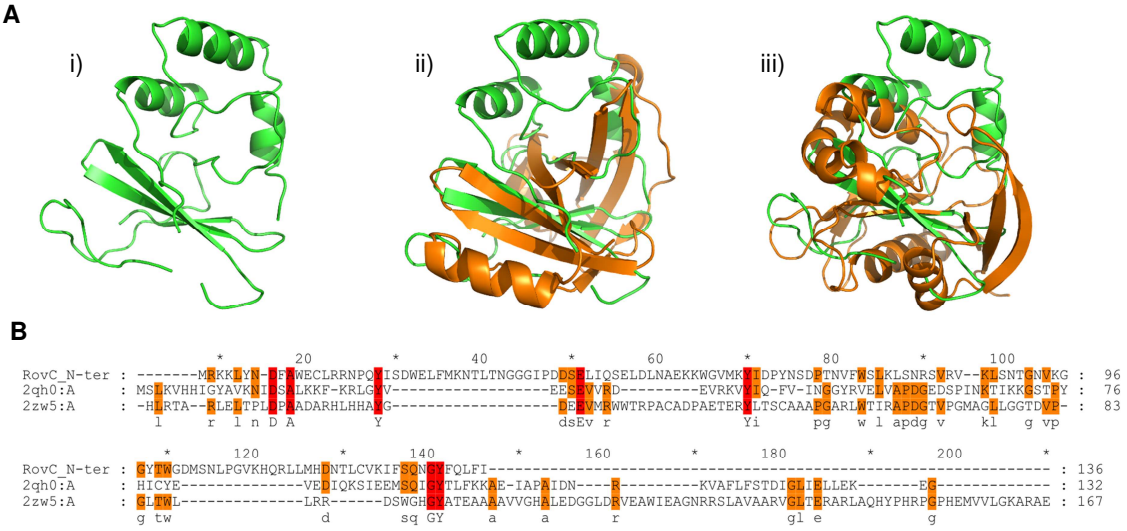

Figure 100: Comparison of the N-terminal domain of RovC with other structurally similar proteins A) i) RovC N-terminal domain (in green) was superimposed with (in orange): ii) lactoylglutathione lyase (PDB code-2qh0:A), iii) Bleomycin acetyltransferase (PDB code-2zw5:A). B) Sequence alignment of RovC N-terminal domain (RovC_N-ter) with lactoylglutathione lyase (2qh0:A) and Bleomycon acetyltransferase (2zw5:A). (100% conserved amino acids are marked in red and 80% in orange). Alignment is generated with GeneDoc (Nicholas et al., 1996).

In contrast, DALI search for the C-terminal domain revealed substantial similarity (Z score>6) with two DNA binding proteins from *Mycobacterium tuberculosis*, RNA polymerase sigma factor (PDB ID-3hug: A) (Thakur et al., 2010) and two component transcriptional regulatory protein (PDB ID-3c57:B) (Wisedchaisri et al., 2008).

The superposition of the C-terminal domain of RovC with these proteins is shown in Figure 101. All of them contain a helix-turn-helix topology suggesting that RovC might be a member of the helix-turn-helix family of DNA binding proteins.

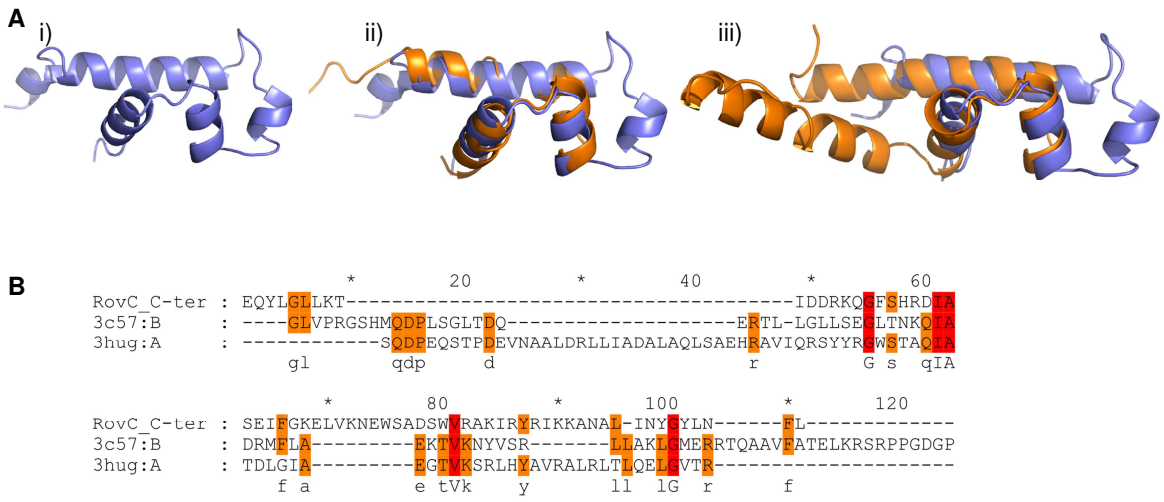

Figure 101: Comparison of the C-terminal domain of RovC with other structurally similar proteins. A) i) RovC C-terminal domain (in blue) was superimposed with (in orange): ii) two component transcriptional regulatory protein (PDB code-3c57:B) (rmsd 2.1 Å), iii) RNA polymerase sigma factor (PDB code-3hug:A) (rmsd 3.4 Å). B) Sequence alignment of RovC C-terminal domain (RovC_C-ter) with two component transcriptional regulatory protein (3c57:B) and RNA polymerase sigma factor (PDB code-3hug:A). (100% conserved amino acids are marked in red and 80% in orange). Alignment is generated with GeneDoc (Nicholas et al., 1996).

5.4.3 RovC-T6SS4 interaction

The DNA binding activity of RovC protein has been characterized *in vitro* by different approaches. Purified RovC showed robust DNA binding activity in EMSA and DNase footprinting assay (Vanessa Knittel, unpublished data). Further with MST, K_d of 11 µM was determined between the 42 bp fragment and RovC (Figure 84). The affinity dropped substantially on reducing the length of the DNA to shorter fragments, indicating that the 38 bp is the minimal DNA binding fragment. Generally, very high affinities (in the nanomolar range) are observed with proteins that target specific DNA sequences (Strauch, 2001). Although the reason for the moderate affinity between RovC and the 42 bp fragment is not known, high salt concentration in the protein buffer (500 mM NaCl) might be the cause (Lohman, 1986; Record et al., 1985). However, high salt was prerequisite for the stability of RovC in solution, so the effect of salt could not be tested on the strength of the interaction.

To determine the stoichiometry of RovC-DNA interaction, stoichiometric titration experiment was performed with MST (as described in (Jerabek-Willemsen et al., 2014)). In contrast to equilibrium titrations, where the concentration of one of the interaction partner is kept well below the K_d, in stoichiometric titrations, concentration of both protein and DNA should be atleast 10-15 fold higher than the K_d to ensure saturation binding. However, for RovC-DNA interaction, for which the K_d of 11 µM was determined, concentration of atleast 100 µM of

labeled DNA was required, which was technically not possible with MST. Therefore, further experiments are required to determine the ratio of DNA to RovC in the complex.

5.4.3.1 DNA binding region on RovC

The conserved basic patch in the C-terminal domain of RovC, forming the outer surface of the oligomeric ring, represents the probable ds DNA binding site (Figure 85). In order to identify if the amino acid residues of this patch are essential for DNA-binding activity, mutational analysis has been performed. Mutants R225, R229, and S219/A220 were found to completely abolish the DNA binding activity and at the same time retained the hexameric structure. On the other hand K56E and K211E/K215E mutants had no influence on DNA binding. Even though K56 is located in the basic patch, the side chain forms a salt bridge with D23 (Figure 102), precluding interaction with the DNA-phosphate backbone. This in turn might explain why the mutation had no effect on DNA binding.

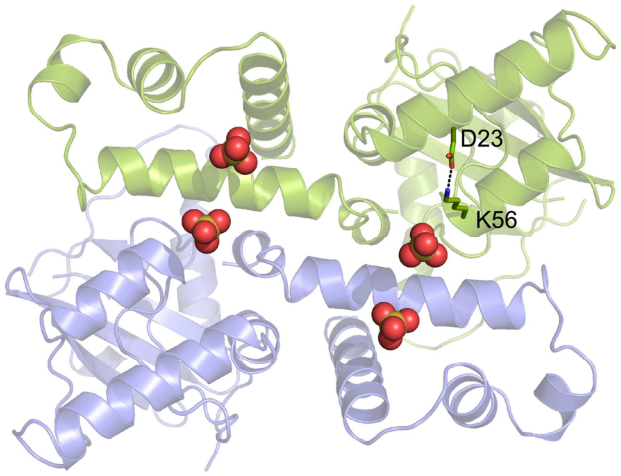

Figure 102: Salt bridge formation in between K56 and D23. Close-up view of the salt bridge formation in between K56 and D23. These two residues are shown as sticks with oxygen atoms in red and nitrogens in blue. Sulfate ions are shown as spheres with sulfur atoms as yellow and oxygen atoms as red.The black dashed line indicates the salt bridge.

RovC C-terminal domain contains a helix-turn-helix motif formed by residues 200-219. This motif is generally known to bind in the major groove of DNA and responsible for making base specific recognition contacts with DNA (Garvie and Wolberger, 2001).

The double mutant K211E/K215E present on the helix-turn-helix motif (Figure 87) had no influence on DNA binding activity indicating that in RovC, this motif is not required for interacting with DNA. Together, these results strongly indicate that the conserved basic patch of dimeric RovC constitutes the DNA binding region. Additionally, it suggests that RovC does not utilize the helix-turn-helix motif for DNA-interaction in the manner observed in classical helix-turn-helix DNA-binding (insertion of one helix into the major groove), but rather as structural element for dimerization and formation of the basic patch.

5.4.4 A model for DNA binding

In the crystal structure, four sulfate ions bind to this basic patch of the dimeric RovC structure (Figure 86). Very often sulfate ions mimic the negatively charged phosphate backbone of DNA.

By taking into account the mutagenesis data, the basic patch and position of the sulfate ions, a 21 bp ds B-DNA molecule was manually docked to the surface of RovC. As seen in Figure 103, the 21 bp DNA fragment binds on the side of the hexameric ring. The DNA fragment fits well in the positively charged patch of RovC and the position of phosphate ions overlaps very well with sulfates present on that patch.

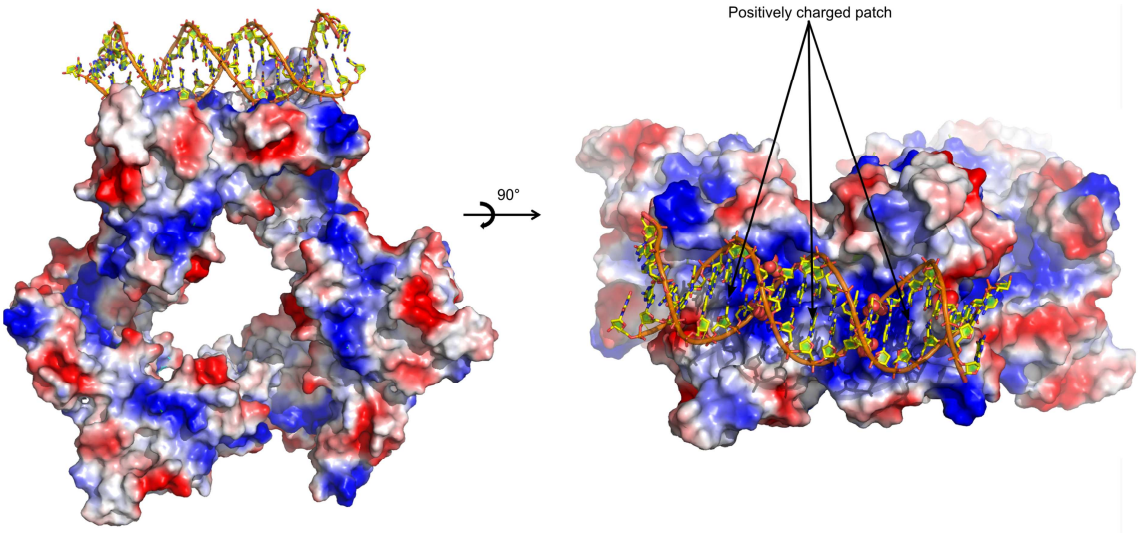

Figure 103: Docked model of protein-DNA complex. Electropotential representation of RovC with the 21 bp B-DNA fragment (PDB ID: 5sy7) docked onto the surface of RovC shown in two different orientations. Potentials between -74.194 kT/e (red) and + 74.194 kT/e (blue) are color-coded. Sulfate ions are shown as spheres with sulfur atoms as yellow and oxygen atoms as red.

Although a 21 bp fragment was docked, but the actual recognition sequence is 38 bp as determined by footprinting assay. Thus, with the current data, a model could be proposed in which hexameric RovC binds to the 38 bp DNA very specifically (Figure 104) and then potentially uses additional sites to bind additional sections of the DNA or a longer fragment of DNA wraps around RovC (Figure 104). However, more data are required for a more detailed prediction.

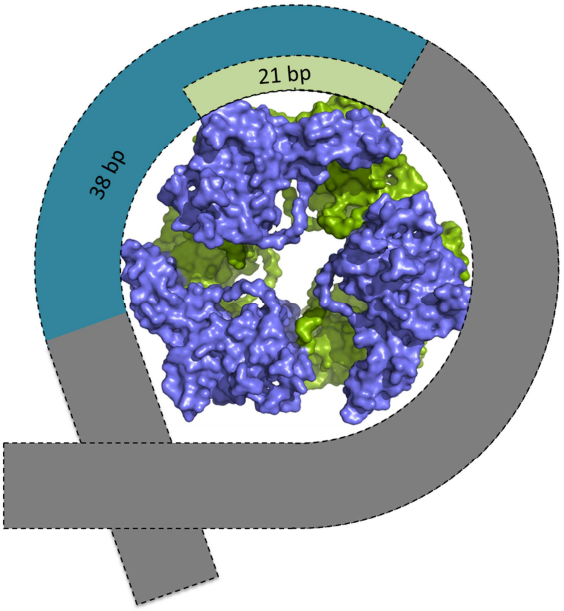

Figure 104: A model for DNA binding. Hexameric RovC (shown as surface) binds to the 38 bp DNA fragment (in blue) very specifically and then might target other regions of DNA (in gray) by using additional sites present on its surface.

5.4.5 Possible mechanisms for transcriptional activation

As already mentioned RovC shares no structural and sequence similarity with any other known transcriptional activator. Hence, the detailed mechanism by which RovC activates transcription remains elusive. Binding of RovC to T6SS4 promoter can result in activation of expression by one of the following mechanisms-

1. By acting as an antirepressor. One example for this is RovA protein. i*nv* gene in *Y. pseudotuberculosis* is repressed by H-NS nucleoid associated protein. Binding of RovA to *inv* promoter causes the displacement of H-NS from the silencing element, leading to expression of Invasin (Lee et al., 2012; Stoebel et al., 2008).
2. By altering the conformation of the promoter. This is seen in case of MerR type activators. Binding of transcriptional activator changes the conformation of the promoter that enables RNA Polymerase to initiate transcription (Browning and Busby, 2004).
3. As a direct transcriptional activator as seen for cAMP receptor protein. Activator binds to the promoter element and initiates transcription by recruiting RNA polymerase to the promoter (Busby and Ebright, 1999).

Which of these mechanisms RovC potentially uses to trigger expression of T6SS, remains an interesting question that needs to be answered in future experiments.

6 Summary and outlook

6.1 InvD

YPIII strain of *Y. pseudotuberculosis* expresses four Invasins, which are InvA, InvB, InvC and InvD (Pisano et al., 2012). So far, details on the regulation, structure and function of InvD have been elusive. In this project, the role of InvD has been elucidated biochemically and structurally. Crystal structure of InvD adhesion domain along with the two Ig-like domains was solved to 2.6 Å resolution. Structural analysis revealed that the adhesion domain adopts a distinct Ig-like fold evolved from its BIg domains. By phage display panning, *in vitro* interaction assays and *ex vivo* experiments, the role of InvD in binding VH3/VK1 fragment of the antibody and B-cells was established. However, the biological significance of the immunoglobulins binding of InvD, the subgroup specificity and interaction with B-cells is presently unknown.

In this context, co-crystallization of InvD with VH3/VK1 antibody will be required to characterize details regarding the InvD-Fab interaction and identify determinants of binding specificity. *In vivo* and cell-culture based experiments are needed to be performed to determine the role of InvD in virulence caused by *Yersinia.*

6.2 InvE

InvE is the newly identified Invasin in the IP31758 strain of *Y. pseudotuberculosis*. The aim of this study was to determine the three dimensional structure of InvE adhesion domain and analyze its function in the context of virulence in *Yersinia*. For that purpose, InvE construct comprising the two N-terminal BIg domains along with the adhesion domain was expressed, purified and crystallized. The three-dimensional structure of InvE was determined to 1.7 Å resolution by SAD and analyzed in solution by SAXS. The two N-terminal domains of InvE adopt a fold which is characteristic of the IgSF fold while the AD shows significant structural resemblance to the CTLDs of InvA and Intimin. Despite their high overall structural similarity, InvE AD does not share any sequence or structural resemblance to the Tir/integrin binding site of Intimin/InvA giving rise to the notion that InvE targets a distinct receptor in the host system.

Possible roles for InvE in immune evasion or colonization remain to be experimentally determined. As Invasins generally target cell surface receptors, binding of InvE to lipids, proteins or carbohydrates should be tested. In addition to that, binding of InvE to different cell lines such as epithelial, T and B cell lines needs to be analyzed. This experiment would give a strong indication of whether InvE targets a cell surface receptor. Once the receptor is identified, it would be interesting to co-crystallize it in complex with InvE to further understand the mechanism of action.

6.3 YadA$_{pstb}$

YadA is a “lollipop” shaped protein on the cell surface of *Y. pseudotuberculosis and Y. pestis* (Hoiczyk et al., 2000). The aim of this project was to provide the structural basis for the functional transition by solving the structure of YadA$_{pstb}$ and to compare it with YadA$_{ent}$ structure (Nummelin et al., 2004). For that purpose, the head domain of YadA$_{pstb}$ was cloned, expressed, purified and crystallized. YadA$_{pstb}$ structure was solved to 2.3 Å resolution. However, detailed information on the architecture of insertion could not be obtained due to high flexibility of that region.

In this context, co-crystallization of the head domain of YadA$_{pstb}$ with fibronectin will shed light on its functional specificity and also stabilize the highly flexible 31 aa region. However, which domain of fibronectin is involved in interaction with YadA$_{pstb}$ is not known. In this context, testing binding of different domains of fibronectin with YadA$_{pstb}$ needs to be performed. In addition, SAXS analysis of the head domain of YadA$_{ent}$ and YadA$_{pstb}$ might provide further insight into differences in the tip region.

6.4 RovC

RovC is a transcriptional regulator of *Yersinia* that activates T6SS4 system (dissertation Stephanie Christine Seekircher, 2014/Vanessa Knittel, unpublished work). The aim of this project was to elucidate the role of RovC biochemically and structurally. For that purpose, full length RovC was over-expressed, purified and crystallized. The three-dimensional structure of RovC was determined to 2.3 Å resolution and analyzed in solution by SAXS. The N-terminal domain of RovC adopts a distinct fold while the C-terminal domain contains a helix-turn-helix motif. SAXS reveals a hexameric ring shaped structure of RovC, which was strongly supported by MALS and site-directed mutagenesis studies. With MST, the affinity between RovC and T6SS4 DNA fragment was determined. Further with structural and mutational analysis, it was shown that R225, R229 and S219/A220 are residues crucial for DNA binding, giving rise to the notion that the DNA binding site is present on the surface of the hexamer. Finally, based on the data obtained, a model was proposed for RovC-mediated DNA-binding in which DNA wraps around the surface of hexameric RovC.

Initial attempts that were made to co-crystallize RovC with T6SS4 DNA were not successful. Future work should focus on the crystallization of RovC in complex with T6SS4 DNA. The co-crystal structure with DNA will provide detailed insights into the mechanism of action of RovC, especially into structural elements involved in the recognition of the specific T6SS-4 DNA-sequence. In this context, testing binding of different fragments of T6SS4 DNA with RovC needs to be performed. Further, SAXS analysis of RovC-DNA complex could also shed light onto the mode of DNA binding. In addition, RovC mutants that have been used in this study should also be analyzed *in vivo*, to analyze their physiological relevance.

7 References

Abdallah, A.M., Gey van Pittius, N.C., Champion, P. a D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M.J.E., Appelmelk, B.J., and Bitter, W. (2007). Type VII secretion–mycobacteria show the way. Nat. Rev. Microbiol. *5*, 883–891.

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, a, and Carniel, E. (1999). Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. U. S. A. *96*, 14043–14048.

Adams, J.J., Nelson, B., and Sidhu, S.S. (2014). Human Monoclonal Antibodies. *1060*, 149–170.

Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Alrutz, M. a., Srivastava, A., Wong, K.W., D'Souza-Schorey, C., Tang, M., Ch'ng, L.E., Snapper, S.B., and Isberg, R.R. (2001). Efficient uptake of Yersinia pseudotuberculosis via integrin receptors involves a Rac1-Arp 2/3 pathway that bypasses N-WASP function. Mol. Microbiol. *42*, 689–703.

Asherie, N. (2004). Protein crystallization and phase diagrams. Methods *34*, 266–272.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. *98*, 10037–10041.

Balada-Llasat, J.-M., and Mecsas, J. (2006). Yersinia Has a Tropism for B and T Cell Zones of Lymph Nodes That Is Independent of the Type III Secretion System. PLoS Pathog. *2*, e86.

Bao, R., Nair, M.K.M., Tang, W., Esser, L., Sadhukhan, A., Holland, R.L., Xia, D., and Schifferli, D.M. (2013). Structural basis for the specific recognition of dual receptors by the homopolymeric pH 6 antigen (Psa) fimbriae of Yersinia pestis. Proc. Natl. Acad. Sci. U. S. A. *110*, 1065–1070.

Barberato, C., Koch, M.H.J., Molecular, E., and Outstation, H. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Crystallogr. *28*, 768–773.

Barlag, B., and Hensel, M. (2015). The giant adhesin SiiE of salmonella enterica. Molecules *20*, 1134–1150.

Barnes, P.D., Bergman, M. a, Mecsas, J., and Isberg, R.R. (2006). Yersinia pseudotuberculosis disseminates directly from a replicating bacterial pool in the intestine. J. Exp. Med. *203*, 1591–1601.

Beaucage, B.Y.G. (1995). Approximations Leading to a Unified Exponential/Power-Law Approach to Small-Angle Scattering. J. Appl. Crystallogr. *8898*, 717–728.

Biedzka-Sarek, M., Jarva, H., Hyytiäinen, H., Meri, S., and Skurnik, M. (2008). Characterization of complement factor H binding to Yersinia enterocolitica serotype O:3. Infect. Immun. *76*, 4100–4109.

Bingle, L.E., Bailey, C.M., and Pallen, M.J. (2008). Type VI secretion: a beginner's guide. Curr. Opin. Microbiol. *11*, 3–8.

Björck, L. (1988). Protein L. A novel bacterial cell wall protein with affinity for Ig L chains. J. Immunol. Baltim. Md 1950 *140*, 1194–1197.

Blocker, A., Komoriya, K., and Aizawa, S.-I. (2003). Type III secretion systems and bacterial flagella: insights into their function from structural similarities. Proc. Natl. Acad. Sci. U. S. A. *100*, 3027–3030.

Bodelón, G., Palomino, C., and Fernández, L.Á. (2013). Immunoglobulin domains in Escherichia coli and other enterobacteria: From pathogenesis to applications in antibody technologies. FEMS Microbiol. Rev. *37*, 204–250.

Bottone, E.J. (1999). Yersinia enterocolitica: overview and epidemiologic correlates. Microbes Infect *1*, 323–333.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. *2*, 57–65.

Brunger, A.T. (1992). Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. Nature *335*, 472–475.

Busby, S., and Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. *293*, 199–213.

Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G.M., Carnemolla, B., Orecchia, P., Zardi, L., and Righetti, P.G. (2004). Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis *25*, 1327–1333.

Casasnovas, J.M., Stehle, T., Liu, J.H., Wang, J.H., and Springer, T. a (1998). A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. Proc. Natl. Acad. Sci. U. S. A. *95*, 4134–4139.

Cascales, E., and Christie, P.J. (2003). The versatile bacterial type IV secretion systems. Nat. Rev. Microbiol. *1*, 137–149.

Chen, X., Kim, T.D., Carman, C.V., Mi, L.-Z., Song, G., and Springer, T.A. (2007). Structural plasticity in Ig superfamily domain 4 of ICAM-1 mediates cell surface dimerization. Proc. Natl. Acad. Sci. U. S. A. *104*, 15358–15363.

Cornelis, G.R. (2002). Yersinia type III secretion: Send in the effectors. J. Cell Biol. *158*, 401–408.

Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., and Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol. *13*, 343–359.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., and Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. Bioinforma. Oxf. Engl. *14*, 892–893.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *62*, 116–124.

Dersch, P., and Isberg, R.R. (2000). An Immunoglobulin Superfamily-Like Domain Unique to the Yersinia pseudotuberculosis Invasin Protein Is Required for Stimulation of Bacterial

Uptake via Integrin Receptors An Immunoglobulin Superfamily-Like Domain Unique to the Yersinia pseudotuberculosis . Infect. Immun. *68*, 2930-2938

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., et al. (2007). In Situ Proteolysis for protein crystallization and structure determination. Nat. Methods. *4*, 1019–1021.

Du, Y., Rosqvist, R., and Forsberg, ke (2002). Role of fraction 1 antigen of Yersinia pestis in inhibition of phagocytosis. Infect. Immun. *70*, 1453–1460.

Duhr, S., and Braun, D. (2006). Why molecules move along a temperature gradient. Proc. Natl. Acad. Sci. U. S. A. *103*, 19678–19682.

Eitel, J., and Dersch, P. (2002). The YadA Protein of Yersinia pseudotuberculosis Mediates High-Efficiency Uptake into Human Cells under Environmental Conditions in Which Invasin Is Repressed. Infect. Immun. *70*, 4880–4891.

Emody, L., Heesemann, J., Wolf-Watz, H., Skurnik, M., Kapperud, G., O'Toole, P., and Wadstrom, T. (1989). Binding to collagen by Yersinia enterocolitica and Yersinia pseudotuberculosis: Evidence for yopA-mediated and chromosomally encoded mechanisms. J. Bacteriol. *171*, 6674–6679.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Eppinger, M., Rosovitz, M.J., Fricke, W.F., Rasko, D. a., Kokorina, G., Fayolle, C., Lindler, L.E., Carniel, E., and Ravel, J. (2007). The complete genome sequence of Yersinia pseudotuberculosis IP31758, the Causative agent of Far East scarlet-like fever. PLoS Genet. *3*, 1508–1523.

Eren, E., Murphy, M., Goguen, J., and van den Berg, B. (2010). An active site water network in the plasminogen activator pla from Yersinia pestis. Struct. Lond. Engl. 1993 *18*, 809–18.

Erhardt, M., and Dersch, P. (2015). Regulatory principles governing Salmonella and Yersinia virulence. Front. Microbiol. *6*, 1–20.

Evans, P., and McCoy, A. (2007). An introduction to molecular replacement. Acta Crystallogr. D Biol. Crystallogr. *64*, 1–10.

Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K., and Stuart, D.I. (1992). Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. Nature *360*, 232–239.

Felek, S., Tsang, T.M., and Krukonis, E.S. (2010). Three Yersinia pestis adhesins facilitate Yop delivery to eukaryotic cells and contribute to plague virulence. Infect. Immun. *78*, 4134–4150.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. *42*, 342–346.

Frenzel, A., Kügler, J., Wilke, S., Schirrmann, T., and Hust, M. (2014). Construction of human antibody gene libraries and selection of antibodies by phage display. Methods Mol. Biol. Clifton NJ *1060*, 215–243.

Galán, J.E., Lara-Tejero, M., Marlovits, T.C., and Wagner, S. (2014). Bacterial Type III Secretion Systems: Specialized Nanomachines for Protein Delivery into Target Cells. Annu. Rev. Microbiol. 415–438.

Garvie, C.W., and Wolberger, C. (2001). Recognition of specific DNA sequences. Mol. Cell *8*, 937–946.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc. Handb. 571–607.

Gjörloff Wingren, A., Hadzic, R., Forsgren, A., and Riesbeck, K. (2002). The novel IgD binding protein from Moraxella catarrhalis induces human B lymphocyte activation and Ig secretion in the presence of Th2 cytokines. J. Immunol. Baltim. Md 1950 *168*, 5582–5588.

Goldschmidt, L., Cooper, D.R., Derewenda, Z.S., and Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Sci. Publ. Protein Soc. *16*, 1569–1576.

Grassl, G. a, Bohn, E., Müller, Y., Bühler, O.T., and Autenrieth, I.B. (2003). Interaction of Yersinia enterocolitica with epithelial cells: invasin beyond invasion. Int. J. Med. Microbiol. IJMM *293*, 41–54.

Griessl, M.H., Schmid, B., Kassler, K., Braunsmann, C., Ritter, R., Barlag, B., Stierhof, Y.-D., Sturm, K.U., Danzer, C., Wagner, C., et al. (2013). Structural insight into the giant Ca^{2+}-binding adhesin SiiE: implications for the adhesion of Salmonella enterica to polarized epithelial cells. Struct. Lond. Engl. 1993 *21*, 741–752.

Griffiths, K., Dolezal, O., Cao, B., Nilsson, S.K., See, H.B., Pfleger, K.D.G., Roche, M., Gorry, P.R., Pow, A., Viduka, K., et al. (2016). i-bodies, Human Single Domain Antibodies That Antagonize Chemokine Receptor CXCR4. J. Biol. Chem. *291*, 12641–12657.

Gueguen, E., Durand, E., Zhang, X.Y., d'Amalric, Q., Journet, L., and Cascales, E. (2013). Expression of a Yersinia pseudotuberculosis Type VI Secretion System Is Responsive to Envelope Stresses through the OmpR Transcriptional Activator. PloS One *8*, e66615.

Hamburger, Z.A., Brown, M.S., Isberg, R.R., and Bjorkman, P.J. (1999). Crystal structure of invasin: a bacterial integrin-binding protein. Science *286*, 291–295.

Hammers, C.M., and Stanley, J.R. (2014). Antibody Phage Display: Technique and Applications. J. Invest. Dermatol. *134*, e17.

Hawgood, B.J. (2008). Alexandre Yersin (1863-1943): discoverer of the plague bacillus, explorer and agronomist. J. Med. Biogr. *16*, 167–172.

Heise, T., and Dersch, P. (2006). Identification of a domain in Yersinia virulence factor YadA that is crucial for extracellular matrix-specific cell adhesion and uptake. Proc. Natl. Acad. Sci. U. S. A. *103*.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C., and Ala'Aldeen, D. (2004). Type V Protein Secretion Pathway : the Autotransporter Story Type V Protein Secretion Pathway : the Autotransporter Story. Microbiol. Mol. Biol. Rev. MMBR *68*, 692–744.

Heroven, A.K., and Dersch, P. (2006). RovM, a novel LysR-type regulator of the virulence activator gene rovA, controls cell invasion, virulence and motility of Yersinia pseudotuberculosis. Mol. Microbiol. *62*, 1469–1483.

Heroven, A.K., and Dersch, P. (2014). Coregulation of host-adapted metabolism and virulence by pathogenic yersiniae. Front. Cell. Infect. Microbiol. *4*.

Heroven, A.K., Nagel, G., Tran, H.J., Parr, S., and Dersch, P. (2004). RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in Yersinia pseudotuberculosis. Mol. Microbiol. *53*, 871–888.

Heroven, A.K., Böhme, K., Rohde, M., and Dersch, P. (2008). A Csr-type regulatory system, including small non-coding RNAs, regulates the global virulence regulator RovA of Yersinia pseudotuberculosis through RovM. Mol. Microbiol. *68*, 1179–1195.

Heroven, A.K., Böhme, K., and Dersch, P. (2012). The Csr/Rsm system of Yersinia and related pathogens: A post-transcriptional strategy for managing virulence. RNA Biol. *9*, 379–391.

Ho, B.T., Dong, T.G., and Mekalanos, J.J. (2014). A view to a kill: the bacterial type VI secretion system. Cell Host Microbe *15*, 9–21.

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., and Heesemann, J. (2000). Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins. EMBO J. *19*, 5989–5999.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, a, Brubaker, R.R., and Garcia, E. (1998). Structural organization of virulence-associated plasmids of Yersinia pestis. J. Bacteriol. *180*, 5192–5202.

Huang, X.Z., Nikolich, M.P., and Lindler, L.E. (2006). Current trends in plague research: From genomics to virulence. Clin. Med. Res. *4*, 189–199.

Isberg, R.R., and Leong, J.M. (1990). Multiple beta 1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. Cell *60*, 861–871.

Isberg, R.R., Voorhis, D.L., and Falkow, S. (1987). Identification of invasin: a protein that allows enteric bacteria to penetrate cultured mammalian cells. Cell *50*, 769–778.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. J. Mol. Struct. *1077*, 101–113.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A., and Finlay, B.B. (1997). Enteropathogenic E. coli (EPEC) transfers its receptor for intimate adherence into mammalian cells. Cell *91*, 511–520.

Kolodziejek, A.M., Schnider, D.R., Rohde, H.N., Wojtowicz, A.J., Bohach, G. a., Minnich, S. a., and Hovde, C.J. (2010). Outer membrane protein X (Ail) contributes to Yersinia pestis virulence in pneumonic plague and its activity is dependent on the lipopolysaccharide core length. Infect. Immun. *78*, 5233–5243.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.

Kozlowski, L.M., Soulika, A.M., Silverman, G.J., Lambris, J.D., and Levinson, A.I. (1996). Complement activation by a B cell superantigen. J. Immunol. Baltim. Md 1950 *157*, 1200–1206.

Krissinel, E. (2015). Stock-based detection of protein oligomeric states in jsPISA. Nucleic Acids Res. *43*, W314–W319.

Kucharzik, T., Lügering, N., Rautenberg, K., Lügering, A., Schmidt, M. a, Stoll, R., and Domschke, W. (2000). Role of M cells in intestinal barrier function. Ann. N. Y. Acad. Sci. *915*, 171–183.

Kügler, J., Wilke, S., Meier, D., Tomszak, F., Frenzel, A., Schirrmann, T., Dübel, S., Garritsen, H., Hock, B., Toleikis, L., et al. (2015). Generation and analysis of the improved human HAL9/10 antibody phage display libraries. BMC Biotechnol. *15*, 10.

Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. *51*, 2778–2786.

Lee, D.J., Minchin, S.D., and Busby, S.J.W. (2012). Activating transcription in bacteria. Annu. Rev. Microbiol. *66*, 125–152.

Leo, J.C., Grin, I., and Linke, D. (2012). Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *367*, 1088–1101.

Leo, J.C., Oberhettinger, P., Schütz, M., and Linke, D. (2014). The inverse autotransporter family: Intimin, invasin and related proteins. Int. J. Med. Microbiol. *305*, 276–282.

Leong, J.M., Fournier, R.S., and Ralph, R. (1990). Identification of the integrin binding domain of the Yersinia pseudotuberculosis invasin protein. EMBO J. *9*, 1979–1989.

Li, B., and Yang, R. (2008). Interaction between Yersinia pestis and the host immune system. Infect. Immun. *76*, 1804–1811.

Lindler, L.E., and Tall, B.D. (1993). Yersinia pestis pH 6 antigen forms fimbriae and is induced by intracellular association with macrophages. Mol. Microbiol. *8*, 311–324.

Lindler, L.E., Klempner, M.S., and Straley, S.C. (1990). Yersinia pestis pH 6 antigen: genetic, biochemical, and virulence characterization of a protein involved in the pathogenesis of bubonic plague. Infect. Immun. *58*, 2569–2577.

Lipovšek, D. (2011). Adnectins: engineered target-binding protein therapeutics. Protein Eng. Des. Sel. *24*, 3–9.

Lohman, T.M. (1986). Kinetics of protein-nucleic acid interactions: use of salt effects to probe mechanisms of interaction. CRC Crit. Rev. Biochem. *19*, 191–245.

Luo, Y., Frey, E.A., Pfuetzner, R.A., Creagh, A.L., Knoechel, D.G., Haynes, C.A., Finlay, B.B., and Strynadka, N.C. (2000). Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex. Nature *405*, 1073–1077.

Makoveichuk, E., Cherepanov, P., Lundberg, S., Forsberg, A., and Olivecrona, G. (2003). pH6 antigen of Yersinia pestis interacts with plasma lipoproteins and cell membranes. J. Lipid Res. *44*, 320–330.

McCoy, A.J. (2006). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. *63*, 32–41.

Mikula, K.M., Kolodziejczyk, R., and Goldman, A. (2012). Yersinia infection tools-characterization of structure and function of adhesins. Front. Cell. Infect. Microbiol. *2*, 169.

Miller, V.L., Beer, K.B., Heusipp, G., Young, B.M., and Wachtel, M.R. (2001). Identification of regions of Ail required for the invasion and serum resistance phenotypes. Mol. Microbiol. *41*, 1053–1062.

Moore, R., and Brubaker, R. (1975). Hybridization of deoxyribonucleotide sequences of Yersinia enterocolitica and other selected members of Enterobacteriaceae. Int. J. Syst. Bacteriol. *25*, 336–339.

Nicholas KB, Nicholas HBJ, Deerfield DWI (1996). GeneDoc: Analysis and visualization of genetic variation, EMBNEW;4: 14.

Nummelin, H., Merckel, M.C., Leo, J.C., Lankinen, H., Skurnik, M., and Goldman, A. (2004). The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. EMBO J. *23*, 701–711.

Oberhettinger, P., Schütz, M., Leo, J.C., Heinz, N., Berger, J., Autenrieth, I.B., and Linke, D. (2012). Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. PloS One *7*, e47069.

Oda, K., Matoba, Y., Noda, M., Kumagai, T., and Sugiyama, M. (2010). Catalytic mechanism of bleomycin N-acetyltransferase proposed on the basis of its crystal structure. J. Biol. Chem. *285*, 1446–1456.

Pakotiprapha, D., and Jeruzalmi, D. (2014). Crystallization of Protein-DNA Complexes. In eLS, John Wiley & Sons Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.

Perry, R.D., and Fetherston, J.D. (2011). Yersiniabactin iron uptake: mechanisms and role in Yersinia pestis pathogenesis. Microbes Infect. *13*, 808–817.

Pierson, D.E., and Falkow, S. (1993). The ail gene of Yersinia enterocolitica has a role in the ability of the organism to survive serum killing. Infect. Immun. *61*, 1846–1852.

Pisano, F., Kochut, A., Uliczka, F., Geyer, R., Stolz, T., Thiermann, T., Rohde, M., and Dersch, P. (2012). In vivo-induced InvA-like autotransporters Ifp and InvC of Yersinia pseudotuberculosis promote interactions with intestinal epithelial cells and contribute to virulence. Infect. Immun. *80*, 1050–1064.

Portnoy, D. a., Moseley, S.L., and Falkow, S. (1981). Characterization of plasmids and plasmid-associated determinants of Yersinia enterocolitica pathogenesis. Infect. Immun. *31*, 775–782.

Rajanna, C., Revazishvili, T., Rashid, M.H., Chubinidze, S., Bakanidze, L., Tsanava, S., Imnadze, P., Bishop-Lilly, K.A., Sozhamannan, S., Gibbons, H.S., et al. (2010). Characterization of pPCP1 plasmids in Yersinia pestis strains isolated from the former Soviet Union. Int. J. Microbiol. *2010*. 1-9

Raman, R., Rajanikanth, V., Palaniappan, R.U.M., Lin, Y.-P., He, H., McDonough, S.P., Sharma, Y., and Chang, Y.-F. (2010). Big domains are novel Ca^2+-binding modules: evidences from big domains of Leptospira immunoglobulin-like (Lig) proteins. PloS One *5*, e14377.

Read, R.J., and McCoy, A.J. (2011). Using SAD data in Phaser. Acta Crystallogr. D Biol. Crystallogr. *67*, 338–344.

Record, M.T., Anderson, C.F., Mills, P., Mossing, M., and Roe, J.H. (1985). Ions as regulators of protein-nucleic acid interactions in vitro and in vivo. Adv. Biophys. *20*, 109–135.

Reis, R.S. Dos, and Horn, F. (2010). Enteropathogenic Escherichia coli, Samonella, Shigella and Yersinia: cellular aspects of host-bacteria interactions in enteric diseases. Gut Pathog. *2*, 8.

Remaut, H., Rose, R.J., Hannan, T.J., Hultgren, S.J., Radford, S.E., Ashcroft, A.E., and Waksman, G. (2006). Donor-Strand Exchange in Chaperone-Assisted Pilus Assembly Proceeds through a Concerted beta Strand Displacement Mechanism. Mol. Cell *22*, 831–842.

Revell, P. a., and Miller, V.L. (2000). A chromosomally encoded regulator is required for expression of the Yersinia enterocolitica inv gene and for virulence. Mol. Microbiol. *35*, 677–685.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. Front. Microbiol. *5*, 172.

Rosqvist, R., Skurnik, M., and Wolf-Watz, H. (1988). Increased virulence of Yersinia pseudotuberculosis by two independent mutations. Nature *334*, 522–524.

Ruffet, E., Pirès, R., Pillot, J., and Bouvet, J.P. (1994). Activation of the classical pathway of complement by non-immune complexes of immunoglobulins with human protein FV (FV fragment-binding protein). Scand. J. Immunol. *40*, 359–362.

Sandkvist, M. (2001). Type II Secretion and Pathogenesis. Infect. Immun. *69*, 3523–3535.

Sasso, E.H., Silverman, G.J., and Mannik, M. (1989). Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. J. Immunol. Baltim. Md 1950 *142*, 2778–2783.

Savin, C., Martin, L., Bouchier, C., Filali, S., Chenau, J., Zhou, Z., Becher, F., Fukushima, H., Thomson, N.R., Scholz, H.C., et al. (2014). The Yersinia pseudotuberculosis complex: Characterization and delineation of a new species, Yersinia wautersii. Int. J. Med. Microbiol. *304*, 452–463.

Schmidt, T.G.M., and Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat. Protoc. *2*, 1528–1535.

Schwarz, S., Hood, R.D., and Mougous, J.D. (2010). What is type VI secretion doing in all those bugs. Trends Microbiol. *18*, 531–537.

Seo, K.S., Kim, J.W., Park, J.Y., Viall, A.K., Minnich, S.S., Rohde, H.N., Schnider, D.R., Lim, S.Y., Hong, J.B., Hinnebusch, B.J., et al. (2012). Role of a new intimin/invasin-like protein in Yersinia pestis virulence. Infect. Immun. *80*, 3559–3569.

Silverman, G.J., and Goodyear, C.S. (2006). Confounding B-cell defences: lessons from a staphylococcal superantigen. Nat. Rev. Immunol. *6*, 465–475.

Simonet, M., Riot, B., Fortineau, N., and Berche, P. (1996). Invasin production by Yersinia pestis is abolished by insertion of an IS200-like element within the inv gene. Infect. Immun. *64*, 375–379.

Snellings, N.J., Popek, M., and Lindler, L.E. (2001). Complete DNA Sequence of Yersinia enterocolitica Serotype 0:8 Low-Calcium-Response Plasmid Reveals a New Virulence Plasmid-Associated Replicon. Infect. Immun. *69*, 4627–4638.

Sodeinde, O.A., Subrahmanyam, Y. V, Stark, K., Quan, T., Bao, Y., and Goguen, J.D. (1992). A surface protease and the invasive character of plague. Science *258*, 1004–1007.

Stoebel, D.M., Free, A., and Dorman, C.J. (2008). Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. Microbiol. Read. Engl. *154*, 2533–2545.

Strauch, M.A. (2001). Protein-DNA Complexes: Specific. In eLS, John Wiley & Sons, Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.

Strong, P.C.R., Hinchliffe, S.J., Patrick, H., Atkinson, S., Champion, O.L., and Wren, B.W. (2011). Identification and characterisation of a novel adhesin Ifp in Yersinia pseudotuberculosis. BMC Microbiol. *11*, 85.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. *25*, 495–503.

Tan, K., Casasnovas, J.M., Liu, J.H., Briskin, M.J., Springer, T. a, and Wang, J.H. (1998). The structure of immunoglobulin superfamily domains 1 and 2 of MAdCAM-1 reveals novel features important for integrin recognition. Struct. Lond. Engl. 1993 *6*, 793–801.

Taylor, G. (2003). The phase problem. Acta Crystallogr. D Biol. Crystallogr. *59*, 1881–1890.

Thakur, K.G., Praveena, T., and Gopal, B. (2010). Structural and biochemical bases for the redox sensitivity of Mycobacterium tuberculosis RslA. J. Mol. Biol. *397*, 1199–1208.

Thornalley, P.J. (2003). Glyoxalase I--structure, function and a critical role in the enzymatic defence against glycation. Biochem. Soc. Trans. *31*, 1343–1348.

von Tils, D., Blädel, I., Schmidt, M.A., and Heusipp, G. (2012). Type II secretion in Yersinia-a secretion system for pathogenicity and environmental fitness. Front. Cell. Infect. Microbiol. *2*, 160.

Torgersen, D., Mullin, N.P., and Drickamer, K. (1998). Mechanism of ligand binding to E- and P-selectin analyzed using selectin/mannose-binding protein chimeras. J. Biol. Chem. *273*, 6254–6261.

Townsley-Fuchs, J., Neshat, M.S., Margolin, D.H., Braun, J., and Goodglick, L. (1997). HIV-1 gp120: a novel viral B cell superantigen. Int. Rev. Immunol. *14*, 325–338.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. Structure *14*, 1617–1622.

Wang, J.-H. (2013). The sequence signature of an Ig-fold. Protein Cell *4*, 569–572.

Weber, B., Hasic, M., Chen, C., Wai, S.N., and Milton, D.L. (2009). Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. Environ. Microbiol. *11*, 3018–3028.

Weis, W.I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W.A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. Science *254*, 1608–1615.

Wisedchaisri, G., Wu, M., Sherman, D.R., and Hol, W.G.J. (2008). Crystal structures of the response regulator DosR from Mycobacterium tuberculosis suggest a helix rearrangement mechanism for phosphorylation activation. J. Mol. Biol. *378*, 227–242.

Yamashita, S., Lukacik, P., Barnard, T.J., Noinaj, N., Tsang, T.M., Krukonis, E.S., Hinnebusch, B.J., and Susan, K. (2012). NIH Public Access. *19*, 1672–1682.

Yang, Y., and Isberg, R.R. (1997). Transcriptional regulation of the Yersinia pseudotuberculosis pH6 antigen adhesin by two envelope-associated components. Mol. Microbiol. *24*, 499–510.

Yang, Y., Merriam, J.J., Mueller, J.P., and Isberg, R.R. (1996). The psa locus is responsible for thermoinducible binding of Yersinia pseudotuberculosis to cultured cells. Infect. Immun. *64*, 2483–2489.

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C.C., Pettersen, E.F., Goddard, T.D., Meng, E.C., Sali, A., and Ferrin, T.E. (2012). UCSF Chimera, MODELLER, and IMP: An integrated modeling system. J. Struct. Biol. *179*, 269–278.

Yi, Y., Ma, Y., Gao, F., Mao, X., Peng, H., Feng, Y., Fan, Z., Wang, G., Guo, G., Yan, J., et al. (2010). Crystal structure of EHEC intimin: insights into the complementarity between EPEC and EHEC. PloS One *5*, e15285.

Zelensky, A.N., and Gready, J.E. (2003). Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). Proteins *52*, 466–477.

Zelensky, A.N., and Gready, J.E. (2005). The C-type lectin-like domain superfamily. FEBS J. *272*, 6179–6217.

Zhang, W., Xu, S., Li, J., Shen, X., Wang, Y., and Yuan, Z. (2011). Modulation of a thermoregulated type VI secretion system by AHL-dependent Quorum Sensing in Yersinia pseudotuberculosis. Arch. Microbiol. 351–363.

Dissertations

Rebecca Geyer (2014). Analysis of the molecular function of invasin-like proteins of *Yersinia pseudotuberculosis* and their role in pathogenesis. Hanover Medical School, Germany.

Stephanie Christine Seekircher (2014). Identification of regulatory factors that control the synthesis of the small regulatory RNA CsrC in *Yersinia pseudotuberculosis*. Technical University Braunschweig, Germany.

8 Appendix

8.1 Abbreviations

List of abbreviations

~	Approximate
∞	Infinite
°C	degree Celsius
µg	Microgram
µl	Microliter
µM	Micromolar
α-C	α-Chymotrypsin
βME	β-mercaptoethanol
aa	amino acids
AE	Actinase
Ail	Adherence and Invasion Locus
AP	alkaline phosphatase
AU	asymmetric unit
BCIP	5-bromo 4-chloro 3'indolylphosphate
Big	bacterial immunoglobulin-like
bp	base pairs
BR	Bromelain
CV	column volume
CL	Clostripain
CTLD	C-type lectin-like domain
Csr	carbon storage regulator
Da	Dalton; 1Da = 1g/mol
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleosid-triphosphate
DTT	Dithiothreitol
dYT	double Yeast Trypton Broth
ECM	extracellular matrix
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EG-C	EndoproteinaseGlu-C
EHEC	enterohemorrhagic *E.coli*
EL	Elastase
EMSA	electrophoretic mobility shift assay
EPEC	enteropathogenic *E.coli*
ESRF	European Synchrotron Radiation Facility
FAE	follicle-associated epithelium
h	Hour

HEp-2	human epithelial type-2
Ifp	Intimin family protein
IgSF	immunoglobulin superfamily
Ilp	Intimin/Invasin-like protein
IM	inner membrane
InvA	Invasin
InvB-E	Invasin(B-E)
IPTG	isopropyl-β-D-thiogalactopyranoside
JCSG	Joint Center for Structural Genomics
K_d	dissociation constant
L	liter
LB	Luria Bertani
M	molar
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
min	minute
ml	mililiter
mM	milimolar
MR-SAD	molecular replacement with single wavelength anomalous dispersion
MS	mass spectrometry
MWCO	molecular weight cut off
NCBI	National Centre for Biotechnology Informatin
Ni-NTA	nickel-nitrilotriacetic acid
nM	nanomolar
OD_{600}	optical density at wavelength of 600 nm
OM	outer membrane
o/n	overnight
PA	Papain
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PE	Pepsin
PK	Proteinase-K
Pla	Plasminogen activator
PMSF	phenylmethanesulfonylfluoride
PVDF	polyvinyl difluoride
pYV	plasmid for *Yersinia* virulence
RFU	relative fluorescence unit
rmsd	root main square deviation
RovC	regulator of virulence associated with CsrC
rpm	revolutions per minute
s	second
S75	superdex 75 (gel filtration column)

S200	superdex 200 (gel filtration column)
SAXS	small angle x-ray scattering
SDS	sodium dodecyl sulftae
SEC-MALS	Size exclusion chromatography with multi angle light scattering
Se-Met	seleno-methionine
SER	Surface Entropy Reduction
SOC	super optimal broth with catabolite repression
SU	Subtilisin
T(1-6)SS	Type (1-6) secretion system
TEMED	tetramethylethylenediamine
TH	Thermolysin
Tris	tris(hydroxymethyl)aminomethane
VH	variable heavy
VL	variable light
(v/v)	volume per volume
(w/v)	weight per volume
Y.	*Yersinia*
YadA	Yersinia adhesin A
Yop	*Yersinia* outer proteins

Abbreviations for amino acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartate/Aspartic acid
E	Glu	Glutamate/Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

8.2 Supplementary table

8.2.1 Glycan binding analysis of InvD

Chart Number	Structure on Masterlist	AvgMeanS-B w/o MIN/MAX	StDev	%CV
1	Gala-Sp8	3	3	92
2	Glca-Sp8	23	26	111
3	Mana-Sp8	8	7	87
4	GalNAca-Sp8	8	4	43
5	GalNAca-Sp15	5	2	51
6	Fuca-Sp8	22	17	79
7	Fuca-Sp9	10	9	89
8	Rhaa-Sp8	17	14	82
9	Neu5Aca-Sp8	4	3	69
10	Neu5Aca-Sp11	20	29	146
11	Neu5Acb-Sp8	29	21	73
12	Galb-Sp8	61	39	64
13	Glcb-Sp8	4	4	124
14	Manb-Sp8	3	2	49
15	GalNAcb-Sp8	30	25	82
16	GlcNAcb-Sp0	7	6	77
17	GlcNAcb-Sp8	9	10	101
18	GlcN(Gc)b-Sp8	1	1	165
19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8	16	3	21
20	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAc-Sp14	13	3	24
21	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8	17	10	62
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	5	3	62
23	6S(3S)Galb1-4GlcNAcb-Sp0	19	6	31
24	(3S)Galb1-4(Fuca1-3)(6S)Glc-Sp0	26	10	39
25	(3S)Galb1-4Glcb-Sp8	4	4	94
26	(3S)Galb1-4(6S)Glcb-Sp0	14	11	81
27	(3S)Galb1-4(6S)Glcb-Sp8	-1	2	-143
28	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp8	3	7	274
29	(3S)Galb1-3GalNAca-Sp8	6	2	38
30	(3S)Galb1-3GlcNAcb-Sp0	13	15	114
31	(3S)Galb1-3GlcNAcb-Sp8	3	8	315
32	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp0	5	4	75
33	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp8	3	3	111
34	(3S)Galb1-4(6S)GlcNAcb-Sp0	28	20	72
35	(3S)Galb1-4(6S)GlcNAcb-Sp8	23	12	51
36	(3S)Galb1-4GlcNAcb-Sp0	12	2	15
37	(3S)Galb1-4GlcNAcb-Sp8	4	4	125
38	(3S)Galb-Sp8	2	2	72
39	(6S)(4S)Galb1-4GlcNAcb-Sp0	0	4	###
40	(4S)Galb1-4GlcNAcb-Sp8	18	32	175
41	(6P)Mana-Sp8	7	13	189
42	(6S)Galb1-4Glcb-Sp0	13	25	184
43	(6S)Galb1-4Glcb-Sp8	7	6	83
44	(6S)Galb1-4GlcNAcb-Sp8	-2	7	-411
45	(6S)Galb1-4(6S)Glcb-Sp8	20	13	65
46	Neu5Aca2-3(6S)Galb1-4GlcNAcb-Sp8	4	6	144
47	(6S)GlcNAcb-Sp8	13	8	58
48	Neu5,9Ac_2a-Sp8	12	25	202
49	Neu5,9Ac2a2-6Galb1-4GlcNAcb-Sp8	7	14	183
50	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	8	98
51	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	8	6	76
52	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	15	65
53	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	7	6	86
54	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	20	23	114

55	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	15	187
56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man-a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	24	21	88
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	24	14	59
58	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9	19	19	104
59	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9	12	8	70
60	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8	3	5	153
61	Fuca1-2Galb1-3GalNAca-Sp8	8	5	63
62	Fuca1-2Galb1-3GalNAca-Sp14	7	5	73
63	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	14	7	50
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp9	18	28	155
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp8	6	3	54
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	19	6	33
67	Fuca1-2Galb1-3GlcNAcb-Sp0	9	8	86
68	Fuca1-2Galb1-3GlcNAcb-Sp8	3	5	152
69	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	11	5	46
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	10	1	11
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0	19	17	87
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8	1	10	653
73	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	2	4	259
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	8	2	24
75	Fuca1-2Galb1-4GlcNAcb-Sp0	12	12	107
76	Fuca1-2Galb1-4GlcNAcb-Sp8	12	12	106
77	Fuca1-2Galb1-4Glcb-Sp0	7	3	42
78	Fuca1-2Galb-Sp8	10	9	92
79	Fuca1-3GlcNAcb-Sp8	9	5	52
80	Fuca1-4GlcNAcb-Sp8	22	18	84
81	Fucb1-3GlcNAcb-Sp8	19	14	75
82	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	10	3	36
83	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	16	22	141
84	(3S)Galb1-4(Fuca1-3)Glcb-Sp0	6	2	36
85	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	4	2	65
86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	6	5	84
87	GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	3	7	204
88	GlcNAcb1-3Galb1-3GalNAca-Sp8	2	3	155
89	GalNAca1-3(Fuca1-2)Galb-Sp8	8	6	78
90	GalNAca1-3(Fuca1-2)Galb-Sp18	39	36	92
91	GalNAca1-3GalNAcb-Sp8	10	9	90
92	GalNAca1-3Galb-Sp8	3	3	102
93	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	17	13	73
94	GalNAcb1-3GalNAca-Sp8	6	1	23
95	GalNAcb1-3(Fuca1-2)Galb-Sp8	26	32	121
96	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	6	10	151
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	10	13	124
98	GalNAcb1-4GlcNAcb-Sp0	10	15	158
99	GalNAcb1-4GlcNAcb-Sp8	31	41	133
100	Gala1-2Galb-Sp8	25	24	95
101	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	5	2	44
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8	5	3	57
103	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	6	2	29
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8	25	27	110
105	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	9	5	53
106	Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0	5	5	102
107	Gala1-3(Fuca1-2)Galb-Sp8	16	20	128
108	Gala1-3(Fuca1-2)Galb-Sp18	30	37	124
109	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8	7	7	101
110	Gala1-3GalNAca-Sp8	29	12	43
111	Gala1-3GalNAca-Sp16	14	14	98
112	Gala1-3GalNAcb-Sp8	9	6	71

113	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	29	14	47
114	Gala1-3Galb1-3GlcNAcb-Sp0	11	6	57
115	Gala1-3Galb1-4GlcNAcb-Sp8	10	6	62
116	Gala1-3Galb1-4Glcb-Sp0	17	19	112
117	Gala1-3Galb1-4Glc-Sp10	2	5	225
118	Gala1-3Galb-Sp8	18	14	82
119	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	5	6	122
120	Gala1-4Galb1-4GlcNAcb-Sp0	5	3	54
121	Gala1-4Galb1-4GlcNAcb-Sp8	4	5	114
122	Gala1-4Galb1-4Glcb-Sp0	11	7	64
123	Gala1-4GlcNAcb-Sp8	9	9	107
124	Gala1-6Glcb-Sp8	7	5	70
125	Galb1-2Galb-Sp8	11	6	55
126	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	21	17	82
127	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	21	21	97
128	Galb1-3(Fuca1-4)GlcNAc-Sp0	19	12	66
129	Galb1-3(Fuca1-4)GlcNAc-Sp8	12	6	54
130	Fuca1-4(Galb1-3)GlcNAcb-Sp8	3	7	208
131	Galb1-4GlcNAcb1-6GalNAca-Sp8	8	10	128
132	Galb1-4GlcNAcb1-6GalNAc-Sp14	9	7	70
133	GlcNAcb1-6(Galb1-3)GalNAca-Sp8	15	26	177
134	GlcNAcb1-6(Galb1-3)GalNAca-Sp14	5	3	59
135	Neu5Aca2-6(Galb1-3)GalNAca-Sp8	10	8	79
136	Neu5Aca2-6(Galb1-3)GalNAca-Sp14	26	3	10
137	Neu5Acb2-6(Galb1-3)GalNAca-Sp8	15	12	76
138	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glcb-Sp10	4	8	178
139	Galb1-3GalNAca-Sp8	35	55	158
140	Galb1-3GalNAca-Sp14	21	8	38
141	Galb1-3GalNAca-Sp16	6	4	74
142	Galb1-3GalNAcb-Sp8	15	13	90
143	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	15	11	72
144	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	19	16	84
145	Galb1-3GalNAcb1-4Galb1-4Glcb-Sp8	35	20	57
146	Galb1-3Galb-Sp8	10	3	33
147	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	15	6	37
148	Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	10	11	115
149	Galb1-3GlcNAcb-Sp0	13	12	90
150	Galb1-3GlcNAcb-Sp8	7	10	141
151	Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	4	156
152	Galb1-4(Fuca1-3)GlcNAcb-Sp8	17	20	123
153	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	15	5	35
154	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	15	6	44
155	Galb1-4(6S)Glcb-Sp0	8	5	65
156	Galb1-4(6S)Glcb-Sp8	13	23	170
157	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	23	22	96
158	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	3	4	117
159	Galb1-4GlcNAcb1-3GalNAca-Sp8	14	9	65
160	Galb1-4GlcNAcb1-3GalNAc-Sp14	23	20	87
161	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	17	14	85
162	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	2	2	102
163	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	10	4	40
164	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	7	4	50
165	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp8	8	7	86
166	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8	16	18	112
167	Galb1-4GlcNAcb1-6(Galb1-3)GalNAc-Sp14	4	6	165
168	Galb1-4GlcNAcb-Sp0	3	4	140
169	Galb1-4GlcNAcb-Sp8	18	9	49
170	Galb1-4GlcNAcb-Sp23	2	3	129
171	Galb1-4Glcb-Sp0	16	25	150
172	Galb1-4Glcb-Sp8	13	6	43

173	GlcNAca1-3Galb1-4GlcNAcb-Sp8	3	8	289
174	GlcNAca1-6Galb1-4GlcNAcb-Sp8	10	9	86
175	GlcNAcb1-2Galb1-3GalNAca-Sp8	35	9	26
176	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8	33	31	94
177	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14	6	9	144
178	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8	3	3	105
179	GlcNAcb1-3GalNAca-Sp8	9	2	17
180	GlcNAcb1-3GalNAca-Sp14	9	5	56
181	GlcNAcb1-3Galb-Sp8	6	12	208
182	GlcNAcb1-3Galb1-4GlcNAcb-Sp0	20	22	109
183	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	10	18	180
184	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	5	2	49
185	GlcNAcb1-3Galb1-4Glcb-Sp0	8	9	108
186	GlcNAcb1-4-MDPLys	41	25	62
187	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8	4	9	229
188	GlcNAcb1-4Galb1-4GlcNAcb-Sp8	33	21	65
189	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	10	9	90
190	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	4	4	103
191	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8	15	11	75
192	GlcNAcb1-6GalNAca-Sp8	22	19	87
193	GlcNAcb1-6GalNAca-Sp14	7	3	48
194	GlcNAcb1-6Galb1-4GlcNAcb-Sp8	42	25	60
195	Glca1-4Glcb-Sp8	1	3	495
196	Glca1-4Glca-Sp8	6	1	21
197	Glca1-6Glca1-6Glcb-Sp8	4	3	65
198	Glcb1-4Glcb-Sp8	3	9	289
199	Glcb1-6Glcb-Sp8	16	8	52
200	G-ol-Sp8	2	4	155
201	GlcAa-Sp8	5	10	213
202	GlcAb-Sp8	7	3	39
203	GlcAb1-3Galb-Sp8	19	14	74
204	GlcAb1-6Galb-Sp8	11	22	192
205	KDNa2-3Galb1-3GlcNAcb-Sp0	-4	8	-210
206	KDNa2-3Galb1-4GlcNAcb-Sp0	7	3	44
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	14	11	75
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	12	2	18
209	Mana1-2Mana1-3Mana-Sp9	12	6	55
210	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	5	4	83
211	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	6	74
212	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	16	11	70
213	Mana1-6(Mana1-3)Mana-Sp9	10	5	51
214	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	10	7	65
215	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	7	8	112
216	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	7	7	105
217	Manb1-4GlcNAcb-Sp0	2	10	434
218	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	16	16	103
219	(3S)Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	16	16	101
220	Fuca1-2(6S)Galb1-4GlcNAcb-Sp0	5	3	48
221	Fuca1-2Galb1-4(6S)GlcNAcb-Sp8	5	12	272
222	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	17	12	69
223	Neu5Aca2-3Galb1-3GalNAca-Sp8	36	15	42
224	Neu5Aca2-3Galb1-3GalNAca-Sp14	15	8	57
225	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	4	4	112
226	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	20	20	100
227	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	11	17	155
228	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	4	5	122
229	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8	13	13	99
230	Neu5Aca2-3(6S)Galb1-4(Fuca1-3)GlcNAcb-Sp8	5	4	72
231	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	6	4	71

232	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8	13	15	120
233	GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	5	7	127
234	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	9	5	58
235	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8	19	25	129
236	Neu5Aca2-3GalNAca-Sp8	30	19	62
237	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0	22	6	29
238	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	7	11	163
239	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8	6	3	46
240	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	27	12	44
241	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	19	12	61
242	Neu5Aca2-3Galb1-3(6S)GalNAca-Sp8	20	12	59
243	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8	3	2	63
244	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	16	22	139
245	Neu5Aca2-3Galb-Sp8	4	3	83
246	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	13	18	146
247	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	3	2	66
248	Fuca1-2(6S)Galb1-4Glcb-Sp0	8	6	73
249	Neu5Aca2-3Galb1-3GlcNAcb-Sp0	10	11	109
250	Neu5Aca2-3Galb1-3GlcNAcb-Sp8	6	4	69
251	Neu5Aca2-3Galb1-4(6S)GlcNAcb-Sp8	9	11	129
252	Neu5Aca2-3Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	17	9	51
253	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	8	18	220
254	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	5	3	58
255	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	5	7	125
256	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8	7	3	43
257	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	14	1	9
258	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	6	6	95
259	Neu5Aca2-3Galb1-4GlcNAcb-Sp0	18	11	60
260	Neu5Aca2-3Galb1-4GlcNAcb-Sp8	35	30	87
261	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	18	11	63
262	Fuca1-2Galb1-4(6S)Glcb-Sp0	15	10	66
263	Neu5Aca2-3Galb1-4Glcb-Sp0	19	20	106
264	Neu5Aca2-3Galb1-4Glcb-Sp8	20	20	97
265	Neu5Aca2-6GalNAca-Sp8	5	4	71
266	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0	0	4	-762
267	Neu5Aca2-6Galb1-4(6S)GlcNAcb-Sp8	19	21	108
268	Neu5Aca2-6Galb1-4GlcNAcb-Sp0	7	7	93
269	Neu5Aca2-6Galb1-4GlcNAcb-Sp8	21	20	95
270	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	10	5	52
271	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	30	14	46
272	Neu5Aca2-6Galb1-4Glcb-Sp0	3	1	31
273	Neu5Aca2-6Galb1-4Glcb-Sp8	19	16	86
274	Neu5Aca2-6Galb-Sp8	10	13	130
275	Neu5Aca2-8Neu5Aca-Sp8	13	5	38
276	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	12	10	81
277	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	6	3	55
278	Neu5Acb2-6GalNAca-Sp8	10	9	84
279	Neu5Acb2-6Galb1-4GlcNAcb-Sp8	12	8	63
280	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	2	5	246
281	Neu5Gca2-3Galb1-3GlcNAcb-Sp0	11	11	99
282	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	2	6	283
283	Neu5Gca2-3Galb1-4GlcNAcb-Sp0	14	11	76
284	Neu5Gca2-3Galb1-4Glcb-Sp0	24	33	141
285	Neu5Gca2-6GalNAca-Sp0	8	18	228
286	Neu5Gca2-6Galb1-4GlcNAcb-Sp0	7	6	85
287	Neu5Gca-Sp8	13	9	64
288	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	18	7	42
289	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	11	4	34
290	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	18	9	50
291	Galb1-4(Fuca1-3)(6S)Glcb-Sp0	30	18	61

292	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	4	3	64
293	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	4	11	259
294	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	4	4	101
295	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	9	6	67
296	4S(3S)Galb1-4GlcNAcb-Sp0	43	42	96
297	(6S)Galb1-4(6S)GlcNAcb-Sp0	4	3	69
298	(6P)Glcb-Sp10	7	2	34
299	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14	29	18	63
300	Galb1-3Galb1-4GlcNAcb-Sp8	18	12	66
301	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	14	15	109
302	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	16	19	123
303	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	22	11	52
304	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	8	6	79
305	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0	10	3	26
306	GalNAcb1-3Galb-Sp8	25	9	37
307	GlcAb1-3GlcNAcb-Sp8	12	10	80
308	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	9	3	27
309	GlcNAcb1-3Man-Sp10	8	11	135
310	GlcNAcb1-4GlcNAcb-Sp10	10	7	71
311	GlcNAcb1-4GlcNAcb-Sp12	31	23	75
312	MurNAcb1-4GlcNAcb-Sp10	14	5	35
313	Mana1-6Manb-Sp10	37	14	37
314	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10	17	19	118
315	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	9	4	48
316	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	8	9	121
317	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	21	27	133
318	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	5	62
319	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	2	8	393
320	Neu5Aca2-8Neu5Acb-Sp17	7	5	72
321	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8	6	5	79
322	Neu5Gcb2-6Galb1-4GlcNAc-Sp8	38	14	36
323	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	22	11	50
324	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	13	13	97
325	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	7	2	22
326	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	7	5	70
327	Neu5,9Ac2a2-3Galb1-4GlcNAcb-Sp0	16	17	106
328	Neu5,9Ac2a2-3Galb1-3GlcNAcb-Sp0	16	17	106
329	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	13	12	93
330	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	39	35	89
331	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	15	22	144
332	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	9	7	78
333	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	6	4	65
334	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	8	11	129
335	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	62	11	19
336	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	16	28	181
337	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	5	7	144
338	GlcNAca1-4Galb1-4GlcNAcb-Sp0	8	2	26
339	GlcNAca1-4Galb1-3GlcNAcb-Sp0	6	3	42
340	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	13	9	65
341	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	70	9	12
342	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	27	27	102
343	GlcNAca1-4Galb1-3GalNAc-Sp14	22	22	103
344	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	5	5	100
345	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	6	4	68
346	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	4	4	89

347	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	1	3	393
348	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	3	4	108
349	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	4	1	33
350	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	41	33	79
351	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	54	33	62
352	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	24	18	77
353	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	28	20	72
354	(6S)GlcNAcb1-3Galb1-4GlcNAcb-Sp0	7	5	75
355	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	13	7	58
356	KDNa2-6Galb1-4GlcNAc-Sp0	14	21	145
357	KDNa2-3Galb1-4Glc-Sp0	-1	3	-220
358	KDNa2-3Galb1-3GalNAca-Sp14	28	25	88
359	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	32	9	28
360	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	39	22	56
361	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAb-Sp20	80	32	40
362	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	15	8	49
363	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	3	8	222
364	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	39	12	30
365	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21	3	5	189
366	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21	9	5	52
367	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	14	16	119
368	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	0	14	###
369	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	15	8	55
370	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	9	6	60
371	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	81	24	30
372	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	14	11	77
373	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	7	5	65
374	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	60	40	67
375	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14	4	1	36
376	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	4	7	163
377	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	6	7	106
378	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	22	21	98
379	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	14	14	95
380	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	3	12	393
381	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp0	5	4	82
382	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	12	18	148
383	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	26	20	78
384	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	5	11	228
385	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	7	6	78
386	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	4	6	146
387	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	12	6	55
388	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	10	7	63
389	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	5	3	70
390	Galb1-3GlcNAcb1-3GalNAca-Sp14	11	6	55
391	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14	61	37	60
392	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	12	8	65
393	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	20	8	41
394	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	36	16	45

395	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	38	29	75
396	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	25	16	63
397	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	5	3	66
398	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14	23	16	68
399	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14	4	3	72
400	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	8	12	159
401	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	14	7	51
402	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	17	17	99
403	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	8	3	41
404	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	14	4	25
405	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0	12	4	31
406	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0	33	48	145
407	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	4	2	47
408	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glcb-Sp21	7	10	152
409	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	16	20	122
410	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	15	7	47
411	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	8	9	118
412	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	2	6	296
413	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	6	3	44
414	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	5	2	54
415	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	57	40	70
416	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	7	6	91
417	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	25	10	39
418	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	14	12	84
419	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	24	18	73
420	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	14	12	88
421	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14	5	3	65
422	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	6	6	91
423	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	12	10	87
424	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14	10	10	102
425	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	48	23	47
426	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	15	5	32
427	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	15	13	87
428	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	10	8	77
429	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	19	8	44
430	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	6	6	100
431	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	4	5	128
432	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	8	7	89
433	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	0	6	###
434	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	3	6	210
435	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	8	9	101
436	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	11	7	59
437	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	10	6	59
438	Galb1-4Galb-Sp10	9	1	7
439	Galb1-6Galb-Sp10	12	4	36
440	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8	5	4	72
441	GalNAcb1-6GalNAcb-Sp8	18	10	55
442	(6S)Galb1-3GlcNAcb-Sp0	16	14	88
443	(6S)Galb1-3(6S)GlcNAc-Sp0	34	36	108

444	Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	42	30	72
445	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	12	10	84
446	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14	6	3	46
447	Galb1-4GlcNAcb1-2Mana-Sp0	11	8	72
448	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAc-Sp14	31	20	65
449	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	44	28	65
450	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	23	18	79
451	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	7	3	43
452	GalNAcb1-4Galb1-4Glcb-Sp0	10	5	48
453	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	34	13	40
454	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	23	8	35
455	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	19	16	81
456	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	38	23	61
457	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	20	18	92
458	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	3	1	43
459	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	15	16	113
460	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	9	9	96
461	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	28	11	41
462	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	9	7	75
463	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	11	5	43
464	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	14	10	74
465	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	11	62
466	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8	16	21	128
467	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8	34	23	67
468	Glca1-6Glca1-6Glca1-6Glcb-Sp10	14	19	136
469	Glca1-4Glca1-4Glca1-4Glcb-Sp10	8	3	42
470	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	17	12	71
471	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	25	10	42
472	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	22	5	21
473	Neu5Aca2-3Galb1-3GlcNAcb1-6(Neu5Aca2-3Galb1-3GlcNAcb1-2)Mana1-6(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	0	5	###
474	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	8	9	115
475	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	10	60
476	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21	3	4	139
477	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0	83	53	64
478	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14	20	10	51
479	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14	14	6	42
480	Neu5Aca2-6Galb1-4 GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14	1	3	181
481	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	38	29	76
482	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	8	2	31
483	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	11	6	56
484	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	17	12	74
485	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	1	4	306

486	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	8	1	19
487	Galb1-3GlcNAcb1-6GalNAca-Sp14	7	8	124
488	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14	9	8	85
489	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14	8	3	30
490	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14	9	8	88
491	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp0	0	5	###
492	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	6	6	100
493	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14	6	2	40
494	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14	9	12	133
495	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	13	6	47
496	Fuca1-2(6S)Galb1-3GlcNAcb-Sp0	17	22	127
497	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14	11	7	61
498	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0	8	6	78
499	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	23	15	63
500	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	14	7	46
501	Neu5Aca2-6GalNAcb1-4(6S)GlcNAcb-Sp8	6	5	79
502	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	6	3	59
503	(3S)GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8	24	13	52
504	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14	21	13	61
505	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14	2	2	138
506	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	5	2	43
507	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	13	19	149
508	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8	11	12	114
509	Galb1-3(6S)GlcNAcb-Sp8	22	14	63
510	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	6	3	44
511	(6S)GalNAcb1-4GlcNAc-Sp8	9	7	72
512	(3S)GalNAcb1-4(3S)GlcNAc-Sp8	79	30	38
513	GalNAcb1-4(6S)GlcNAc-Sp8	30	27	87
514	(3S)GalNAcb1-4GlcNAc-Sp8	21	23	107
515	(4S)GalNAcb-Sp10	8	1	9
516	Galb1-4(6P)GlcNAcb-Sp0	4	8	192
517	(6P)Galb1-4GlcNAcb-SP0	9	5	49
518	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAc-Sp14	12	12	99
519	Neu5Aca2-6Galb1-4GlcNAcb1-2Man-Sp0	18	26	139
520	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0	5	2	42
521	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0	3	2	73
522	GalNAca1-3(Fuca1-2)Galb1-4 GlcNAcb1-2Mana-Sp0	11	10	86
523	Galb1-3GlcNAcb1-2Mana-Sp0	13	8	60
524	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAc-Sp14	44	28	62
525	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0	14	6	41
526	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0	7	6	82
527	GalNAcb1-4GlcNAcb1-2Mana-Sp0	10	13	127
528	Neu5Aca2-3Galb1-3GalNAcb1-4Galb1-4Glcb-Sp0	12	12	96
529	GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	6	11	169
530	Galb1-4GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	2	3	123
531	Galb1-4GlcNAcb1-2 Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	11	7	63
532	Fuca1-4(Galb1-3)GlcNAcb1-2 Mana-Sp0	9	9	108
533	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	6	6	97
534	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAc-Sp0	1	2	211
535	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	25	28	112
536	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	25	8	33
537	Galb1-3GalNAcb1-3Gal-Sp21	9	9	96
538	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	14	13	93
539	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	38	9	24
540	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-	1017	152	15

	2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12			
541	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	37	6	16
542	Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	28	48	172
543	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	9	9	105
544	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	551	41	7
545	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	41	8	19
546	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	175	95	54
547	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	21	8	37
548	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	35	21	59
549	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	27	21	77
550	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp25	6	3	62
551	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	3	4	166
552	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	6	8	137
553	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	2	5	198
554	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0	25	18	70
555	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAc-Sp0	16	7	44
556	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	8	6	71
557	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Man a1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	14	11	73
558	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAc-Sp24	39	40	102
559	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	125	67	53
560	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14	19	24	126
561	GalNAcb1-3GlcNAcb-Sp0	9	6	75
562	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0	28	20	72
563	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	11	6	50
564	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	10	10	100
565	GlcNAb1-3Galb1-3GalNAc-Sp14	22	13	61
566	Galb1-3GlcNAcb1-6(Galb1-3)GalNAc-Sp14	20	13	64
567	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	19	6	34
568	(3S)GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	6	7	111
569	(3S)GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0	15	10	65
570	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	53	12	23
571	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	29	19	66
572	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp21	14	14	105
573	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	13	6	47
574	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	71	28	39
575	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	114	52	46
576	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	47	13	29
577	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-	53	10	19

	3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24			
578	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	34	11	33
579	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	20	14	71
580	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	18	12	65
581	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	101	10	10
582	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	264	27	10
583	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	122	29	23
584	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	70	6	8
585	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	49	30	61
586	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	14	7	53
587	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	53	60	113
588	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	107	97	91
589	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	89	9	10
590	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	104	36	35
591	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	10	7	68
592	GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	11	12	106
593	GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	10	7	74
594	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	6	2	39
595	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	11	5	48
596	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	12	8	68
597	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	8	14	180
598	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14	20	16	77
599	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	21	26	122
600	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	21	15	71
601	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	58	48	83
602	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	26	2	9
603	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	15	11	73
604	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	42	25	61
605	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	69	108	155
606	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	42	6	14
607	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	0	3	###
608	GlcNAcb1-3Fuca-Sp21	6	2	27
609	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp21	30	26	87

Chart number indicates the glycan number; **Structure on masterlist** indicates the name of the glycan; **AvgMeanS-B w/o MIN/MAX** indicates the average RFU value from the replicates; **St Dev** indicates the Standard deviation; **%CV** is 100 * St.Dev/Mean.

Acknowledgements

At this point, I would like to thank several people who contributed to the success of this work.

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Andrea Scrima for his support, guidance, constant willingness to discuss and for his trust in me during my work in the past three and a half years. I am very thankful to him for giving me intellectual freedom in my work, opportunities to present my research work in various conferences and for his constructive feedback and invaluable suggestions during my thesis.

I would like to thank Prof. Dr. Wulf Blankenfeldt and Prof. Dr. Stefan Dübel for agreeing to be on my examination committee and Prof. Dr. Michael Hust for serving as the head of my Ph.D. defense commission.

I would like to acknowledge helpful suggestions from my thesis committee members: Dr. Joop van den Heuvel and Prof. Dr. Petra Dersch.

My sincere thanks goes to all those who have collaborated in this work. I would like to show my gratitude to Prof. Dr. Petra Dersch for collaboration in InvD and RovC projects, helpful discussions and giving me access to her laboratory and research facilities. I thank Rebecca Geyer for performing *in vivo* studies on InvD, Vanessa Knittel for performing EMSAs and footprinting assays and Dr. Ann Kathrin Heroven for her suggestions. I thank Carina Schmühl for her tremendous help and suggestions in microbiology experiments.

I thank Jörn Pezoldt and Prof. Dr. Jochen Hühn for contributing their analysis on InvD binding to B-cells and for their suggestions. I thank Prof. Dr. Michael Hust and Saskia Helmsing for performing panning experiments, analyzing the data and producing antibodies in mammalian cells. I thank Prof. Dr. Teresa Carlomagno and Dr. Luca Codutti for their help with SEC-MALS experiments.

I extend my thanks to Dr. Justyna Sikorska for her helpful suggestions, friendship and performing STD-NMR experiments. I also thank Prof. Dr. Carlo Unverzagt for providing the glycan. I thank Dr. Joop van den Heuvel and Maren Bleckmann for producing antibodies in mammalian cells and to all other members of RPEX for their help.

I would like to acknowledge the funding of the glycan array screening experiment and the participation of the Protein-Glycan Interaction Resource of the CFG and the National Center for Functional Glycomics (NCFG) at Emory University.

I would like to thank all my fellow and past group members. It was a great pleasure working with this group. Many thanks to Dr. Stefan Schmelz for his help and suggestions on working with graphics programs and in solving structures, Dr. Caroline Behrens for her continuous support and extensive discussions, Petra Völler for organizing everything in the lab, Brendan Willmann for performing test expressions when I was busy in writing my thesis and Dr.

Manfred Wirth for his help in microscopy. I also thank Stefan Leupold, Milica Bajagic, and Archna for providing a healthy work atmosphere.

Furthermore, I would like to thank all the members of SFPR especially Maurice Diwo, Florian Witzgall and Monika Popp for their extensive discussions on scientific problems and good humor and friendship. My special appreciation goes to Nina Schwemmlein for reading parts of my thesis, her friendship and encouragement. I also thank Dr. Jan Pippel for his help with Pymol and helpful suggestions. Thanks to Dr. Joern Krausze for making me familiar to SAXS data processing.

My deepest appreciation goes to my parents for being my strength, their undying belief in me and endless support; and to my sister, wherever you are, you will continue to be my inspiration.

Curriculum vitae

Name: Pooja Sadana

Date of birth: 30.09.1990

Place of birth: Saharanpur, India

ACADEMIC CAREER

07/2013-present **PhD student** at the Helmholtz Centre for Infection Research, Braunschweig, Germany.

Department- Structure and Function of Proteins. Group of Dr. Andrea Scrima.

- Structural and functional characterization of adhesins and the virulence regulator of *Yersinia pseudotuberculosis.*

07/2011-05/2013 **M.Sc. Biotechnology** at the Indian Institute of Technology, Roorkee, India.

Department of Biotechnology. Group of Dr. Shailly Tomar.

- Cloning, expression, purification and crystallization of nonstructural protein1 from Chikungunya virus.

Final grade: 9.3 (on the scale of 1(lowest) to 10 (highest))

07/2008-05/2011 **B.Sc. (Hons) Chemistry** at Hindu College, University of Delhi, India.

Final marks: 88.44%

AWARDS/SCHOLARSHIPS

03/2015 **Poster Award**, 1st prize for poster presentation, 23rd Annual Conference of the German Crystallographic society, Göttingen, Germany.

07/2011-05/2013 **Department of Biotechnology scholarship** from the government of India.

02/2012 **M.K. Rastogi award**, for standing first in Hindu College in B.Sc. (Hons) Chemistry 3rd year.

01/2012 **S.P. Suri memorial award**, for standing first in the University of Delhi in B.Sc. (Hons) Chemistry 3rd year.

PRESENTATIONS

12/2016 **Oral presentation**, 9th International PhD symposium of the Helmholtz International Graduate School, Braunschweig, Germany.

09/2016 **Oral presentation**, 19th Heart-of-Europe Bio-Crystallography meeting, Burg Warberg, Germany.

12/2015 **Poster presentation**, 8th International PhD symposium of the Helmholtz International Graduate School, Braunschweig, Germany.

03/2015 **Poster presentation**, 23rd Annual Conference of the German Crystallographic society, Göttingen, Germany.

12/2014 **Poster presentation**, 7th International PhD symposium of the Helmholtz International Graduate School, Braunschweig, Germany.

10/2014 **Poster presentation**, North Regio Day on Infection V, HZI, Braunschweig, Germany.

09/2014 **Oral presentation**, 17th Heart-of-Europe Bio-Crystallography meeting, Berlin, Germany.

05/2014 **Poster presentation**, 5th Annual Retreat, Helmholtz Centre for Infection Research Graduate School, Goslar Hahnenklee, Germany.